现代生物技术前沿

张 今 编著

进化生物技术

——酶定向分子进化

35

科学出版社
www.sciencep.com

中科院植物所图书馆
S0003774
58.17425
552

现代生物技术前沿

张 今 编著

进化生物技术
——酶定向分子进化

科 学 出 版 社
北 京

28131

内 容 简 介

本书较为全面地介绍了酶定向分子进化原理、策略、方法、应用以及发展趋势。主要内容包括定向分子进化思想、概念和原理；定向分子进化的基本策略和方法；定向分子进化的筛选/选择方法；核酶和脱氧核酶定向分子进化；蛋白质酶定向分子进化；外显子改组定向进化酶；抗体酶定向分子进化；蛋白质组、代谢途径、病毒和细胞的定向进化。

本书可供从事生物化学、分子生物学、生物工程学、化学和药学等相关专业的科研人员和教学工作者参考，也适用于综合性大学、医学和农业院校相关学科的高年级本科生和研究生阅读。

图书在版编目(CIP)数据

进化生物技术：酶定向分子进化/张今编著. —北京：科学出版社，2004
(现代生物技术前沿)
ISBN 7-03-012639-4

Ⅰ.进… Ⅱ.张… Ⅲ.酶-定向分子-分子进化 Ⅳ.Q55

中国版本图书馆 CIP 数据核字(2003)第 120917 号

责任编辑：马学海 李久进 贾学文/责任校对：柏连海/封面设计：王浩
排版制作：科学出版社编务公司/责任印制：安春生

科学出版社 出版
北京东黄城根北街16号
邮政编码：100717
http://www.sciencep.com
新蕾印刷厂 印刷
科学出版社发行 各地新华书店经销
*
2004年4月第 一 版 开本：B5(720×1000)
2004年4月第一次印刷 印张：16 1/2
印数：1－3 000 字数：309 000

定价：38.00元
(如有印装质量问题，我社负责调换〈新欣〉)

序

酶定向分子进化旨在模拟突变、重组和选择的自然进化机制，体外进化新酶。定向分子进化实质上是达尔文进化论在分子水平上的延伸和应用，是一个新兴的技术科学领域。过去10年,特别是近两年,定向分子进化在生物分子结构与功能研究方面,在工业、农业和医药等领域呈现强劲的发展势头。这表明进化论不再仅仅是一种理论，而是逐渐发展为一门应用科学——“进化生物技术”。进化生物技术关系人类命运(例如，HIV是世界上最积极进化的生物体，如何用进化生物技术制服它；如何用进化生物技术战胜细菌耐药性；如何用进化生物技术解决生物品种灭亡等重大问题)，定向分子进化则是进化生物技术中的关键技术。

张今教授编著的《进化生物技术——酶定向分子进化》一书对酶定向分子进化原则、策略和方法，现状和发展趋势进行了比较全面的介绍。他在广泛收集国内外资料的基础上，结合自己多年的实践经验，使本书内容具有新颖性和实用性，对读者颇有参考价值和启迪作用。我相信《进化生物技术——酶定向分子进化》一书的出版将会推动我国进化生物技术的发展。

孙曼霁

前　言

定向分子进化(directed molecular evolution)又称实验室进化(laboratory evolution)或进化生物技术(evolutionary biotechnology)。它旨在体外模拟自然进化机制(突变、重组和选择),使进化过程朝着人们需要的方向发展。定向分子进化的实质是达尔文进化论思想在分子水平上的延伸和应用,用进化和组合技术研究和解决复杂生物学和化学问题已成为化学和生物学迅猛发展的领域之一。

定向分子进化的思想是1967年由Spiegelman及其同事提出的,他们当时定向进化的目标是RNA(Spiegelman S et al. Proc Natl Acad Sci USA, 1967, 58: 217 ~ 224)。1984年Eigen和Gardiner提出了分子进化的理论(Eigen M, Gardiner W. Pure Appl Chem, 1984, 56: 967 ~ 978)。1993年Arnold研究组首先应用定向分子进化的原理改进酶,提出定向分子进化的方法——易错PCR(error-prone PCR, epPCR)(Chen K, Arnold F H. Proc Natl Acad Sci USA, 1993, 90: 5618 ~ 5622)。1994年Stemmer发展了一种方法,称之为DNA改组(DNA shuffling),有力地推动了定向分子进化的发展(Stemmer W P. Nature, 1994, 370: 389 ~ 391)。现在定向进化已由基因和蛋白质分子向基因组和蛋白质组、代谢途径和病毒,甚至向细胞定向进化方面努力延伸,细胞的定向分化和定向进化可能是生物工程的未来。

近10年,定向分子进化在生物分子结构与功能研究方面,在工业、农业和医药等领域已展示出广阔的应用前景,高通量筛选是发现新酶的重要工具。可以预见,在不久的将来,定向分子进化的发展将使生物催化剂及催化过程以崭新的面貌出现在生命科学和生物技术“世界”。

本书的意图是向读者介绍定向分子进化原理、策略、方法、应用以及发展趋势,以期促进定向分子进化在我国的发展。本书1 ~ 7章围绕核酸和蛋白质定向分子进化展开,第8章介绍蛋白质组、代谢途径、病毒和细胞定向进化。本书编写有两个指导原则,一个是讨论一般原理和采用特殊实例并举的原则;另一个是讨论学科交叉和技术集成在定向分子进化领域的现状。每章末所引参考文献尽量选择最近发表的文章和能提供更广泛的文献目录的文献,便于读者进一步阅读。在编写过程中,力求在有限篇幅中尽可能介绍基本原理,同时也力求最大限度反映出近年来本领域的进展。在内容取舍方面,既注意到避免与相关著作内容重复,又尽可能自成系统。本书从基因和蛋白质→基因组和蛋白质组→代谢途径、病毒和细胞,结合自己的实践经验(本课题组自1990年至今主要从事酶定向分子进化方面的研究工作),比较详细地介绍了酶定向分子进化的理论和实践。

由于本人学术水平有限，错误和欠妥之处在所难免，诚恳地希望广大读者批评指正，不胜企盼之至。

在本书编写过程中，美国西南医学中心张雷博士、加州大学孔祥铎博士、麻省理工学院(MIT)刘文胜博士提供了有效的文献资料；书稿的电脑输入、图表制作和校对由孙妍红工程师、李爽博士(参加第2章和第3章部分编写)和李宾硕士完成，特此致谢！

中国科学院院士孙曼霁先生为本书作序，在此表示深深的谢意！

吉林大学分子酶学工程教育部重点实验室

张　今

2003年8月18日

目　录

序
前言
第 1 章　定向分子进化的思想、概念和原理 ……………………………………（1）
1.1　引言 ……………………………………………………………………（1）
1.2　体内进化 ………………………………………………………………（2）
1.3　体外进化 ………………………………………………………………（4）
1.4　*in silico* 进化 …………………………………………………………（8）
参考文献…………………………………………………………………………（9）
第 2 章　定向分子进化的基本策略和方法 ……………………………………（11）
2.1　引言……………………………………………………………………（11）
2.2　基本策略………………………………………………………………（11）
2.3　基本方法………………………………………………………………（12）
2.3.1　易错 PCR …………………………………………………………（12）
2.3.2　DNA 改组和族改组 ………………………………………………（13）
2.3.3　外显子改组 ………………………………………………………（16）
2.3.4　合成改组 …………………………………………………………（16）
2.3.5　体外随机引发重组 ………………………………………………（17）
2.3.6　交错延伸 …………………………………………………………（18）
2.3.7　酶法体外随机-定位诱变 …………………………………………（19）
2.4　非同源基因改组………………………………………………………（19）
2.4.1　过渡模板随机嵌合 ………………………………………………（19）
2.4.2　增长截短 …………………………………………………………（20）
2.4.3　非顺序同源蛋白质重组…………………………………………（22）
2.4.4　合理设计与随机组合方法………………………………………（22）
2.4.5　模块组装 …………………………………………………………（24）
2.4.6　任意数量的碱基随机插入和缺失 ………………………………（26）
2.4.7　片段突变 …………………………………………………………（28）
2.4.8　设计的寡核苷酸组装 ……………………………………………（28）
2.4.9　体外区室化 ………………………………………………………（31）
参考文献 ………………………………………………………………………（31）

第3章 定向分子进化的选择/筛选策略和方法 …… (34)
3.1 引言 …… (34)
3.2 核糖体和 mRNA 展示技术 …… (36)
3.3 噬菌体展示技术 …… (39)
3.3.1 噬菌体展示作为酶库筛选的工具 …… (39)
3.3.2 噬菌体-酶 …… (43)
3.3.3 噬菌体-酶的选择 …… (44)
3.4 酵母表面展示系统 …… (57)
3.5 酵母 n-杂交系统 …… (59)
3.5.1 酵母双杂交检验 …… (59)
3.5.2 酵母双杂交系统 …… (61)
3.6 细菌双杂交(B2H)系统 …… (66)
3.7 蛋白质互补检验 …… (67)
3.8 酵母 n-杂交系统的应用 …… (68)
3.8.1 蛋白质-蛋白质相互作用 …… (68)
3.8.2 蛋白质-DNA 相互作用 …… (68)
3.8.3 蛋白质-RNA 相互作用 …… (68)
3.8.4 蛋白质-小分子相互作用 …… (68)
3.9 环境库的构建和筛选 …… (69)
参考文献 …… (73)
第4章 核酶和脱氧核酶定向分子进化 …… (76)
4.1 引言 …… (76)
4.2 核酶 …… (81)
4.2.1 锤头型核酶 …… (81)
4.2.2 发夹型核酶 …… (84)
4.2.3 蛋白质-RNA 复合酶(RNaseP) …… (84)
4.2.4 组 I 内含子和组 II 内含子 …… (86)
4.2.5 环状核酶 …… (87)
4.3 脱氧核酶 …… (89)
4.3.1 10-23 结构脱氧核酶 …… (90)
4.3.2 8-17 结构脱氧核酶 …… (90)
4.3.3 手枪型结构脱氧核酶 …… (91)
4.3.4 “二分”型结构脱氧核酶 …… (91)
4.3.5 环状结构脱氧核酶 …… (92)
4.4 核酶和脱氧核酶的体外进化 …… (93)

4.4.1 拓展核酸顺序空间的新方法 …… (93)
4.4.2 体外选择别构核酶 …… (97)
4.4.3 核酶的连续体外进化 …… (101)
4.4.4 核蛋白酶的体外选择 …… (102)
4.4.5 延伸蛋白质编码的核酶体外选择 …… (104)
4.4.6 体外选择环状核酶 …… (107)
4.4.7 核酶/脱氧核酶的催化潜能和进化策略 …… (110)
参考文献 …… (113)
第 5 章 蛋白质酶定向分子进化 …… (118)
5.1 引言 …… (118)
5.2 自然界蛋白质进化机制 …… (119)
5.2.1 基因复制 …… (119)
5.2.2 串联复制 …… (120)
5.2.3 环状变换 …… (121)
5.2.4 寡聚合 …… (122)
5.2.5 基因融合 …… (122)
5.2.6 结构域募集 …… (123)
5.2.7 外显子改组 …… (124)
5.3 定向分子进化 …… (125)
5.3.1 组合定向进化 …… (125)
5.3.2 模块定向进化 …… (130)
5.3.3 杂合酶 …… (140)
5.3.4 定向共进化 …… (145)
5.3.5 适应性进化 …… (146)
5.3.6 稳定性和活性的进化 …… (150)
5.3.7 底物特异性的进化 …… (151)
5.3.8 对映体选择性的进化 …… (155)
5.3.9 $(\beta/\alpha)_8$-桶酶的功能进化 …… (170)
5.4 应用进化策略研究酶的结构和功能 …… (172)
5.4.1 应用进化策略研究分支酸变位酶的结构和功能 …… (172)
5.4.2 应用进化策略研究 DNA 聚合酶的结构和功能 …… (178)
参考文献 …… (184)
第 6 章 外显子改组定向进化酶 …… (186)
6.1 引言 …… (186)
6.2 外显子改组的分子机制 …… (187)

6.3 蛋白质结构域是运动的功能模块 …………………………… (189)
6.4 外显子改组产生模块酶 …………………………… (192)
6.4.1 血液凝集和血纤维溶解的蛋白酶 …………………………… (192)
6.4.2 模块交换的原理 …………………………… (196)
6.4.3 模块交换的机制 …………………………… (198)
6.5 体外外显子改组 …………………………… (199)
6.5.1 策略和方法 …………………………… (199)
6.5.2 组合结构域重组 …………………………… (201)
6.5.3 体外外显子改组实例 …………………………… (203)
参考文献 …………………………… (204)
第7章 抗体酶定向分子进化 …………………………… (206)
7.1 引言 …………………………… (206)
7.2 抗体酶定向分子进化策略 …………………………… (207)
7.2.1 突变策略 …………………………… (207)
7.2.2 组合抗体库策略 …………………………… (208)
7.2.3 噬菌体抗体库 …………………………… (209)
7.2.4 合成抗体库策略 …………………………… (209)
7.2.5 结合荧光素 scFv 抗体库 …………………………… (210)
7.2.6 T-细胞受体策略 …………………………… (210)
7.2.7 提高催化活性策略 …………………………… (211)
7.3 抗体酶定向分子进化筛选/选择策略 …………………………… (211)
7.3.1 结合、催化和免疫组分策略 …………………………… (211)
7.3.2 理论检验策略 …………………………… (212)
7.4 抗体酶定向分子进化筛选/选择方法 …………………………… (216)
7.4.1 直接筛选 …………………………… (216)
7.4.2 化学筛选 …………………………… (216)
7.4.3 遗传选择 …………………………… (217)
7.4.4 动力学选择 …………………………… (219)
7.4.5 应用反应化合物体外选择抗体酶 …………………………… (221)
7.4.6 催化抗-DNA 抗体的筛选方法 …………………………… (222)
7.4.7 选择性感染噬菌体 …………………………… (223)
参考文献 …………………………… (224)
第8章 蛋白质组、代谢途径、病毒和细胞的定向进化 …………………………… (225)
8.1 引言 …………………………… (225)
8.2 结构基因组定向进化的策略和方法 …………………………… (226)

8.2.1 融合报告 …………………………………………………… (227)
8.2.2 宿主细胞应激反应报告 ………………………………………… (229)
8.2.3 直接检验 ………………………………………………………… (229)
8.3 基因组改组 ………………………………………………………… (230)
8.3.1 基因组改组改进泰乐菌素生产菌表型 …………………… (231)
8.3.2 基因组改组改进乳酸生产菌表型………………………………… (235)
8.4 生物合成途径定向进化 …………………………………………… (237)
8.4.1 β-葡糖苷酸酶定向进化 ……………………………………… (237)
8.4.2 三嗪水解酶类定向进化 ……………………………………… (239)
8.4.3 类胡萝卜素生物合成途径定向进化 …………………………… (240)
8.4.4 L-甲硫氨酸生物合成途径定向进化 …………………………… (243)
8.5 病毒定向进化 …………………………………………………… (245)
8.6 微生物细胞定向进化 …………………………………………… (248)
参考文献…………………………………………………………………… (250)

第 1 章　定向分子进化的思想、概念和原理

1.1 引　　言

过去 10 年,定向进化(directed evolution)在生物分子结构与功能研究方面及工业、农业、医药等领域的应用呈现出强劲的发展势头[1]。特别是近两年,定向进化研究的论文以令人惊奇的数量和速度增加。

关于生物进化的新颖思路可以追溯到《物种起源》一书[2]。在《物种起源》问世140 多年中,进化理论已经有了很大发展。现代生物学不断地证实达尔文的惊人想像,尽管这位伟人根本不懂遗传学和分子生物学。

关于生物进化的学说可以围绕进化动力、进化方向、进化速度等 3 个方面归纳如下[3]:

- 进化动力
  - 外因论——布丰学说、莱伊尔学说、新拉马克主义、米丘林-李森科主义、新灾变论
  - 内因论——经典拉马克主义、活力论、终极目的论、突变论、某些现代分子进化学说
  - 内因-外因论(遗传突变 + 选择)——达尔文学说、现代综合论
- 进化方向
  - 非定向
    - 随机论——莱伊尔学说、分子进化中性论
    - 局部环境适应论——达尔文学说、现代综合论
  - 定向——新旧拉马克主义、活力论、终极目的论、现代“环境趋向变化论”
- 进化速度
  - 渐变——莱伊尔学说、达尔文学说、现代综合论
  - 跳跃——断续平衡论、新灾变论
  - 恒定——分子进化中性论

早期的进化学说基本倾向于决定论;近几十年的新学说基本倾向于随机论;近几年进化学说基本倾向于基因驱动(gene driver)说和新灾变论(新决定论)。100 多年来,进化论的发展似乎在决定论和随机论之间摆动, 达尔文学说似乎介于两者之间。

在分子水平上体外定向进化生物分子的思想是 1967 年 Spiegelman 及其同事[4]提出的,当时他们定向进化的对象是 RNA。1984 年 Eigen[5]及 1993 年 Kauffman[6]提出分子进化的理论。1991 年我们建立了“随机-定位诱变”定向进化的方

法[7,8]，又被称为 REM(random and extensive mutagenesis)法[9,10]。1993 年 Arnold 研究组[11]应用分子进化的原理创造性地改进酶，提出易错 PCR(epPCR)。1994 年 Stemmer[12]发展了一种方法，称之为 DNA 改组(DNA shuffling)。1999 年 Stemmer 等[13]把 DNA 改组延伸到组改组(family shuffling)，扩大了顺序空间(sequence space)，进而提出分子育种(molecular breeding)。1998 年在美国加利福尼亚州召开的“IBC 第二届工业酶定向进化讨论会”，以及第 14～16 届国际酶工程会议都将定向进化列为会议专题。现在定向进化从基因和蛋白质→基因组和蛋白质组→代谢途径和病毒甚至细胞方向发展。

定向进化的实质是达尔文进化论在分子水平上的延伸和应用[1]。定向进化(也称实验室进化或进化生物技术或定向分子进化)是在体外模拟突变、重组和选择的自然进化机制，使进化朝着人们需要的方向发展。定向进化虽然人工模拟自然进化，但两者却不相同：①进化动力不同。同源蛋白质的比较指出，保守突变是自然进化的驱动力[14]，而定向进化中非保守取代也可明显改进蛋白质的功能[15]；②进化方向不同。定向进化是人为制造特殊条件，模拟自然进化机制，使进化过程定向进行，主要是蛋白质突变的适应积累。定向进化的“适应度”(fitness，也称遗传漂移，drift)主要指生物分子的稳定性和活性。中性突变的组合可以产生新功能的嵌合蛋白质，称为“结构中性进化”[16](construction neutral evolution)或“选择中性”[17](selective neutrality)。自然进化是机体的适应突变(adaptive mutation)和中性突变(neutral mutation)的积累。自然进化的“适应度”主要指机体生存和繁殖的能力。 ③进化速度不同。定向进化只需几年、几月、几周、甚至几天的时间，而自然进化是自发出现的一非常缓慢的过程，需几百万年。自然选择使进化向有利于生物体适应生存和繁殖环境的方向发展。环境和适应方式的多样性决定进化方向的多样性。④定向进化的目标往往超越生物学意义的要求，以寻求自然环境中并不需要的功能，所有可能的氨基酸顺序，即顺序的多样性(X 个氨基酸链长的蛋白质顺序有 X^{20} 种)组成蛋白质的顺序空间。所有可能的顺序空间组成“适应度景观”(fitness landscape)。对于定向进化而言，将研究每一个性质或性质组合的变化，探索所希望的蛋白质在顺序空间的可及性[17～19]。

1.2 体 内 进 化[20]

群体遗传学微分方程一般是为有性复制物种导出的，这种思想基本上以重组作为变异的主要来源，突变则被认为是特殊的事件。在生物多聚物进化设计中，与上述观点正相反，突变是变异的共同来源，重组则是特殊事件，例如基因改组。在排除无性情况下，通过微分适应度讨论选择的数学表达式如下：

$$\frac{dX_k}{dt} = X_k\left(f_k - \sum_{j=1}^{n} f_j X_j\right) = X_k(f_k - \Phi);k = 1,2,\cdots,n \tag{1}$$

式中，X_k 表示突变体的分数；f_k 是适应值；Φ 作为群体的平均适应值。方程(1)直接解释为：当突变体的 $f_k - \Phi$ 是正或它的适应值是在平均值以上，即 $f_k > \Phi$，那么 dX_k/dt 则为正，并且这个突变体频率增加；相反，如果 $f_k < \Phi$，那么相应突变体的分数将减少，并且最终接近零，突变体消失。这样选择具有最大适应值 $f_{max} = \max\{f_k,\ k = 1,2,\cdots,n\}$ 的突变体，并且在足够长时间之后，惟有该突变体将在群体中存在，$\lim_{t\to\infty} X_{max} = 1$。换句话说，如果时间相当长，所有不太适应的突变体将消失，群体将变为均一。当新种由于偶然事件处于消失的危险之中，选择过程和结果只是由适应值的差所决定：$S = f_\lambda - f_{max}$，S 值是通过选择优势突变体所需要的世代数来反映(图 1.1)。在自然选择中，优势突变体一般很少出现，新突变体在可以被群体中接收之前需要进化上千代。

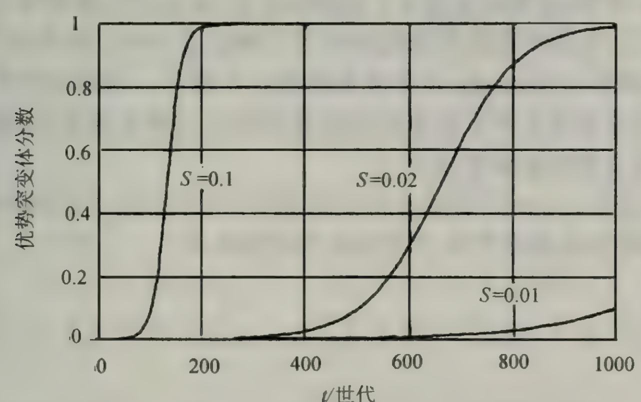


图 1.1　优势突变体的选择示意图

曲线表明根据方程 $X(t) = X_0/\{X_0 + (1 - X_0)\exp(-St)\}$，在 $N = 10\ 000$ 个体的群体中突变体的选择和固定。时间 t 是以世代数或复制步骤计量，X_0 是群体中新突变体的最初频率，$S = f' - f$ 是它的选择优势。曲线表明在群体中采用单拷贝的初始条件，$X = 0.0001$

群体遗传学基本是 Motoo Kimura 思想的延伸。Kimura 指出适应突变是相当少的，大多数突变体在选择上是中性的，进化的主要作用是消除有害突变体。Kimura 观点是由蛋白质和核酸顺序比较分析资料支持的，这成为目前分子系统发育的基础。基因型是稳步地改变，并且当表型停滞时基因型也在稳步地改变。尽管分子资料以压倒的优势支持中性进化，但是第一个直接证据还是最近细菌进化

的实验:表型性质的改变,基因组 DNA 顺序也继续改变[21];当表型停滞时,基因组 DNA 顺序以同样速度或者甚至在适应周期时改变更快[22]。Kimura 途径是以选择过程随机描述为基础的:每个新形成的突变体达到一定的固定的概率(随 S 值增加),它是相对于群体中当时优势类型的适应度。在中性情况下,$S=0$,群体通过随机漫步方式在顺序空间移动。随机漫步使基因型连续空间扩大(它与顺序空间概念相似)。 20 世纪 90 年代,中性进化的计算机模拟确定了 Kimura 观点。

1.3 体外进化[4,18]

在试管中模拟自然进化的思想是 20 世纪 60 年代由 Spiegelman 及其同事在哥伦比亚大学提出的[4]。他们的实验如图 1.2 所示。病毒 RNA 分子被转移到含有适于病毒 RNA 复制介质(Qβ复制酶及核苷三磷酸单体)的试管中,产生修饰的子代。70 次转移,RNA 合成的速率大约增加一个数量级(图 1.2)。这就是体外定向进化的经典实验。

在 Watson 和 Crick 揭示核酸分子的结构之后,进化的各种群体的空间性质是清楚的。顺序空间是指突变或重组的所有 DNA(或 RNA)顺序的分离空间(图 1.3)。顺序空间是以 Hamming 距离度量的超立方体[23]。顺序空间有两个非常重要的性质:①它是具有各种长度基因组的高维物,λ 以核苷酸度量;②顺序空间的所有的点,即所有顺序是相等的。

关于分子进化,Eigen[24]综合分子生物学和化学反应动力学的知识,提出通过动力学方程网络方式描述复制、突变和选择的模型,即

$$\frac{dX_k}{dt} = X_k(Q_{kk}f_k - \Phi) + \sum_{j=1,j\neq k}^{n} Q_{kj}f_jX_j;\ k = 1,2,\cdots,n \tag{2}$$

式中,Q_{kk}表示 $I_k \rightarrow 2I_k$ 的概率,Q_{kj}表示突变体($I_k \rightarrow I_k + I_j$)的概率;$\Phi$ 与适应度权重相同。复制过程是一种产生正确拷贝产物平行的各种反应网络。方程(2)可以提供一种研究分子进化的实验方案。非中性复制集合的最适基因型称为主顺序(master sequence)(图 1.4)。如果突变是一种相当复杂的事件,主顺序会被密切相关或不相关的突变体云(mutant cloud)所围绕。在适当条件下,主顺序及它的突变体云接近稳态的基因型分布,称为准种(quasi-species)。准种概念的提出,对于了解进化是重要的。

突变频率只取决于由 I_k 到 I_j 突变必须交换的单体数,它通过 Hamming 距离计算 $d(I_j, I_k)$,有

$$Q_{kj} = q^{\lambda}\varepsilon^{d(I_j,I_k)} = q^{\lambda}\left(\frac{1-q}{q}\right)^{d(I_j,I_k)} \tag{3}$$

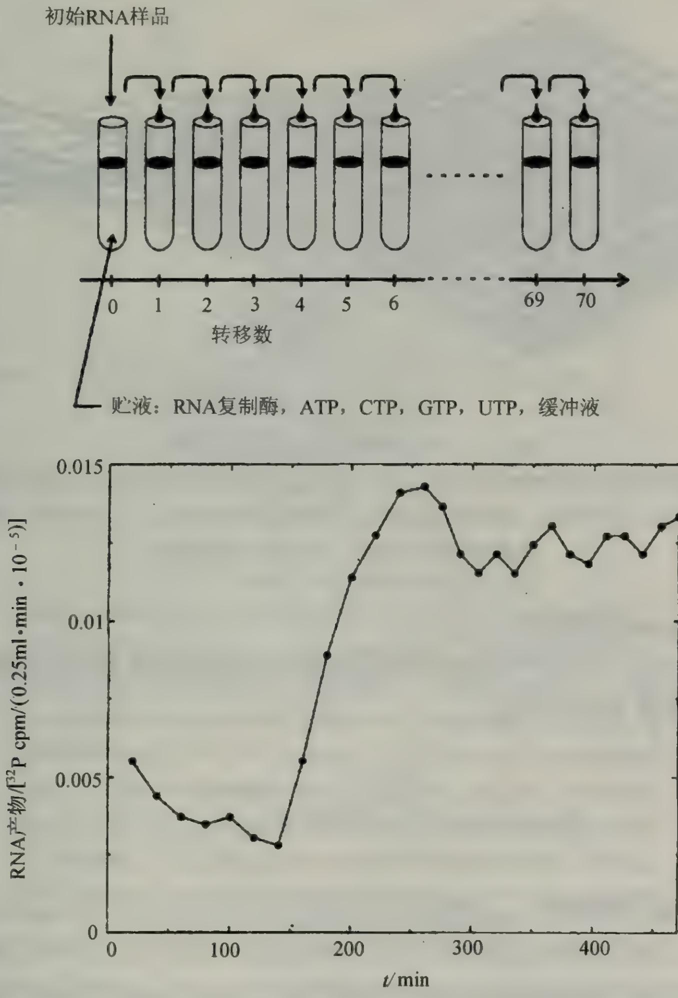


图 1.2　RNA 进化实验图

图上部分表示应用系列转移技术在试管中进化 RNA 分子。在一定的时间间隔之后，样品被转移到下一个含有新鲜贮液的试管中，补充 RNA 合成时所消耗的材料。贮液含有复制所需要的酶（例如 Qβ-复制酶）和活化的单体（ATP，UTP，GTP 和 CTP）。图下部分表示 RNA 合成速率。它是通过放射性 GTP 移入新产生的 RNA 分子来测量。复制速率表明逐步增加。早期降低是由于通过主顺序首先形成的准种适应度变低[4]

式中，独立的参数 $\varepsilon = (1-q)/q$ 是突变率和精确度之间的比率。在这个模型中，突变速率可以由 3 个量来表达，聚合物链长 λ，复制的精确度 q（通常以 $p=1-q$

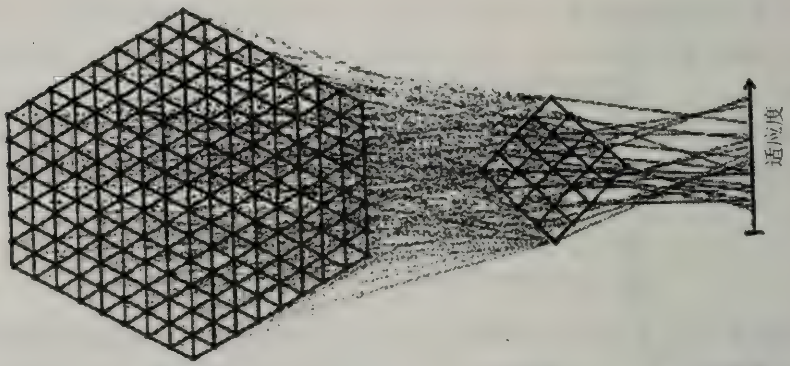


图 1.3　顺序空间和基因型-表型图解

该图表明顺序(或基因型)空间到表型空间，并进一步转为实际数。第二图是“适应度”景观，它可以通过三维标绘图来说明。两图通常是由多到少，是不可逆的。顺序和表型空间是高维的,这里是用二维说明的

表示)和 Hamming 距离 $d(I_j, I_k)$。

方程(3)可以为突变平衡表征的稳态提供相当精确的复制过程,这种复制的最小精确度是容易直接估计的。

$$Q_{\min} = q^{\lambda}_{\min} = \frac{\sum_{k=1, k \neq m}^{n} X_k f_k / (1 - X_m)}{f_m} \tag{4}$$

式中,复制的最小精确度 $q_{\min}$相当于最大耐受突变率,$P_{\max} = 1 - q_{\min}$,它被称为错误临界值。在突变速率高于临界值时,准种被相同分布取代。换句话说,包括主顺序的所有突变体以同样概率出现。当复制精确度太低时,对生物多聚物来讲,相同分布不可能存在。通过顺序空间群体的随机漂移,这种高度错误倾向再现现象可以看作随机复制。

准种的概念用病毒群体和计算机模拟已证实[25]。假定两个高适应度的基因型,每个由特殊的突变体云所围绕(图 1.4),基因型 I_{m1}同 I_{m2}比较有较高的适应度,但是在具有较低平均适应度的突变体云中,准种可以看作是突变速率 P 的函数,临界复制精确度 $q_{cr} = 1 - P_{cr}$。在低突变速率 $P > P_{cr}$时,适应值的差异决定选择,因此具有较高适应度的主顺序 I_{m1}占优势。然而,上述临界突变率 $P > P_{cr}$,突变回流到主顺序是决定性的,然后选择 I_{m2}。

分子准种概念是建立在变异和选择基础上的一类新生物技术平台的基础。关

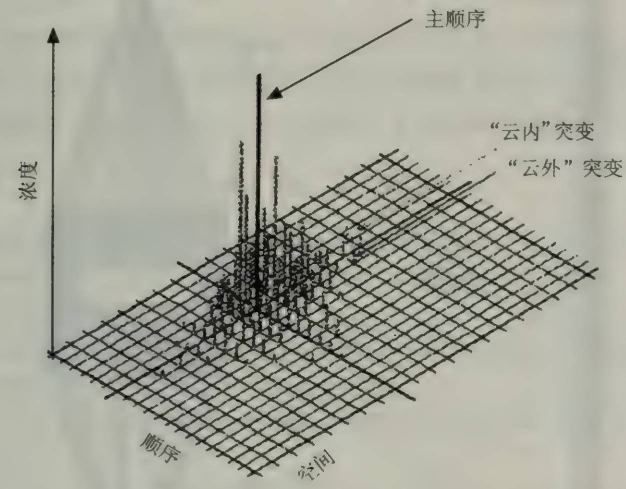


图 1.4 围绕主顺序准种型突变体分布

准种是一种在顺序空间中多核苷酸顺序(RNA 或 DNA)的有序分布。最适应的基因型或主顺序 I_m，一般代表最高频率，并被紧密相关顺序"云"(cloud)围绕。顺序相关是以突变体数表示。 在点突变的情况下，顺序之间的距离是 Hamming 距离。准种被定义为方程(2)稳定不变的解。复制错误速率在最大值之下称为错误阈，这样才有固定解。在临界错误速率之上，方程(2)的固定解是相同的，均一分布。然而，均一分布在自然界或体外从未实现。因为可能的核苷酸顺序(4^{λ}, λ是链长)超过个体数几个数量级。由不正确复制所决定的实际性质导致随机漂移，群体通过顺序空间移动，产生两类突变:"云内"突变(灰色)，产生早已存在的突变体;"云外"突变(黑色)，产生新基因型

于进化生物技术的结果见文献[25,26]。进化设计的主要思想是由连续选择循环组成的，每一循环包括三相:①扩增;②多样化;③选择(图 1.5)。核酸分子的扩增和多样化现在已成为分子生物学常规的方法，适体(aptamer)的产生就是一个例子。配体进化设计的通用的方法是 SELEX(通过指数富集配体的系统进化)技术。一般，20~30 个选择循环足以获得最佳结合分子。在有利的情况下，可能获得结合常数在 nmol/L 范围的适体[28]。另一个例子是核酶(ribozyme)，已经选择出各种新的核酶(见第 4 章)。

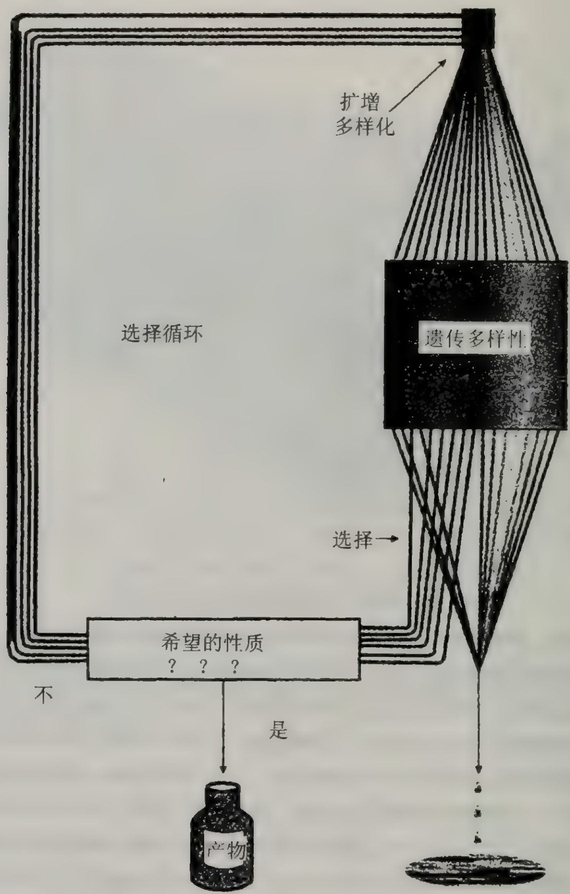


图 1.5 在选择循环中生物多聚物的进化设计

生物分子的性质，例如与靶结合或催化功能，是通过选择循环最佳化。每个循环由三相组成：①扩增；②多样化；③选择。扩增和多样化由分子生物学方法实现，例如 PCR 和自身持续顺序复制反应(3SR 反应)。两者都可提高突变率。选择需要敏感的方法，例如 SELEX 和化学标签

1.4 *in silico* 进化

许多不同的基因型可以给出同样的表型，并且取决于条件，即基因型 + 环境型 = 表型，不同的表型可以享有同样的适应值。一个以上顺序可以导致同样的结构。在控制条件下中性进化的直接证据是用大肠杆菌超过万代系列转移实验证实[28]。基于 RNA 分子性质的模型已经建立：基因型是用 RNA 顺序表示，表型是用 RNA 二

级结构模型表示。这样，最小自由能的二级结构折叠的 RNA 顺序决定 RNA 基因型和表型之间的关系。

一般考察基因型-表型图谱（图 1.3）可获得如下结果：①顺序数（基因型）超过二级结构（表型）大约几个数量级；②共同结构少，稀有结构多；③每个结构可能含有一个以上的顺序。上述结论是 *in silico* 进化（主要是神经网络法）得出的[29]。

参 考 文 献

1 田波，李传昭，孙伦泉，王学. 分子进化工程. 北京：科学出版社，1999

2 达尔文. 物种起源. 周建人，叶笃庄，方宗熙译. 北京：商务印书馆，1991

3 张昀. 生物进化. 北京：北京大学出版社，1998

4 Mills D R, Peterson R L, Spiegelman S. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. Proc Natl Acad Sci USA, 1967, 58: 217 ~ 224

5 Eigen M, Gardiner W. Evolutionary molecular engineering based on RNA replication. Pure Appl Chem, 1984, 56: 967 ~ 978

6 Kauffman S. The origins of order: self-organization and selection in evolution. New York: Oxford University Press, 1993

7 张红缨，李正强，张今. 酶法体外建立天冬氨酸酶的随机突变库. 科学通报，1991，19：1500 ~ 1502

8 Zhang H Y, Lin L, Du W Y, *et al*. Enhancement of stability and activity of aspartase by random and site-directed mutagenesis. Biochem Biophys Res Commun, 1993, 192: 15 ~ 21

9 Ling M M, Robinson B M. Approach to DNA mutagenesis: an overview. Anal Biochem, 1997, 254: 157 ~ 178

10 Moffet D A, Hecht M H. De novo proteins from combinatorial libraries. Chem Rev, 2001, 101: 3191 ~ 3203

11 Chen K, Arnold F H. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. Proc Natl Acad Sci USA, 1993, 90: 5618 ~ 5622

12 Stemmer W P. Rapid evolution of a protein *in vitro* by DNA shuffling. Nature, 1994, 370: 389 ~ 391

13 Stemmer W P, Soon N M. Molecular breeding of viruses for targeting and other clinical properties. Tumor Targeting, 1999, 4: 59 ~ 62

14 Go M, Miyazawa S. Relationship between mutability, polarity and exteriority of amino acid residues in protein evolution. Int J pept Protein Res, 1980, 15: 211 ~ 224

15 Miyazaki K, Arnold F H. Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. J Mol Evol, 1999, 49: 716 ~ 720

16 Stoltzfus A. On the possibility of constructive neutral evolution. J Mol Evol, 1999, 49: 169 ~ 181

17 Demetrius L. Selective neutrality and enzyme kinetics. J Mol Evol, 1997, 45: 370 ~ 377

18 Voigt C A, Kauffman S, Wang Z G. Rational evolutionary design: the theory of *in vitro* protein evolution. Adv Protein Chem, 2001, 55: 79 ~ 160

19 Brakmann S, Johnsson K. Directed molecular evolution of proteins. Wiley-VCH Verlag Gnnb H, Weinheim, 2002

20 Papadopoulos D, Schneider D, Meier-Eiss J M, *et al*. Genomic evolution during a 10 000generation experiment with bacteria. Proc Natl Acad Sci USA, 1999, 96: 3807 ~ 3812

21 Huynen M A, Stadler P F, Fontana, W. Smoothness within ruggedness: the role of neutrality in adaptation. Proc Natl Acad Sci USA, 1996, 93: 397 ~ 401

22 Hamming R W. Coding and information theory. Prentice Hall. 2nd ed. NJ: Englewood Cliffs, 1989

23 Eigen M. Selforganization of matter and the evolution of biological macromolecules. Naturwis- senschaften, 1971, 58: 465 ~ 523

24 Wilke C O, Wang J I, Ofria C, *et al*. Evolution of digital organisms at high mutation rates leads to survival of the flattest. Nature, 2001, 412: 331 ~ 333

25 Drake W, Charlesworth B, Charlesworth D, *et al*. Rates of spontaneous mutation. Genetics, 1998, 148: 1667 ~ 1686

26 Burch C L, Chao L. Evolvability of an RNA virus is determined by its mutational neighbourhood. Nature, 2000, 406: 625 ~ 628

27 Wilson D S, Szostak J W. *In vitro* selection of functional nucleic acids. Ann Rev Biochem, 1999, 68: 611 ~ 647

28 Elena S F, Cooper V S, Lenski R E. Punctuated evolution caused by selection of rare beneficial mutations. Science, 1996, 272: 1802 ~ 1804

29 Babajide A, Farber R, Hofacker I L, *et al*. Exploring protein sequence space using knowledge-based potentials. J Theor Biol, 2001, 212: 35 ~ 40

第 2 章　定向分子进化的基本策略和方法

2.1　引　　言

应用定向进化和组合技术研究和解决复杂的生物学和化学问题,已成为化学和生物学中最活跃的领域之一。

定向分子进化的第一个例子可追溯到 Spiegelman 和 Eigen 及 Gardiner 等经典实验。他们的实验证实进化途径适合工程生物分子。近 10 年定向分子进化途径的成功已使其成为进化生物技术的关键技术。这不仅取决于策略和方法的效力,而且更取决于解决问题的能力和范围,它可以解决合理途径所不能解决的问题,因而应用面越来越广。

我们可以把定向分子进化看作突变加筛选/选择的重复循环。定向分子进化的第一步是产生多样性。随机突变策略是产生多样性的最基本的策略。最直接的方法是通过易错 PCR(epPCR)或饱和突变把随机突变引入目的基因。此外,通过 DNA 改组或相关技术产生附加的多样性,迅速积累有益突变。同源基因之间的改组,也称族改组,是一种有效的方法,但是扩展顺序空间也是有限的。于是科学家开发出非同源基因之间产生嵌合体的各种策略和方法,进一步拓展顺序空间,本章做了重点介绍。本章所介绍的策略和方法不可能涵盖定向分子进化的所有方面。充其量说,本章也只能向读者提供定向分子进化策略和方法的“快照”。

2.2　基本策略

定向分子进化的基本策略如图 2.1 所示[1]。第一轮随机突变中所选择的突变体再重复进行随机突变和选择,这样会积累更多的有益突变[图 2.1(a)]。应用 DNA 改组或其他方法改进突变体的重组,可使有益突变组合并消除有害突变[图 2.1(b)]。与单一基因突变体的随机突变和重组[图 2.1(a)、(b)]相反,当同源基因重组产生嵌合体时,可以出现新功能[图 2.1(c)]。与随机突变不同,同源基因重组所积累的氨基酸改变主要是中性突变,大部分形成功能突变体。这就是说中性突变的组合形成的嵌合蛋白质可以产生新的功能。

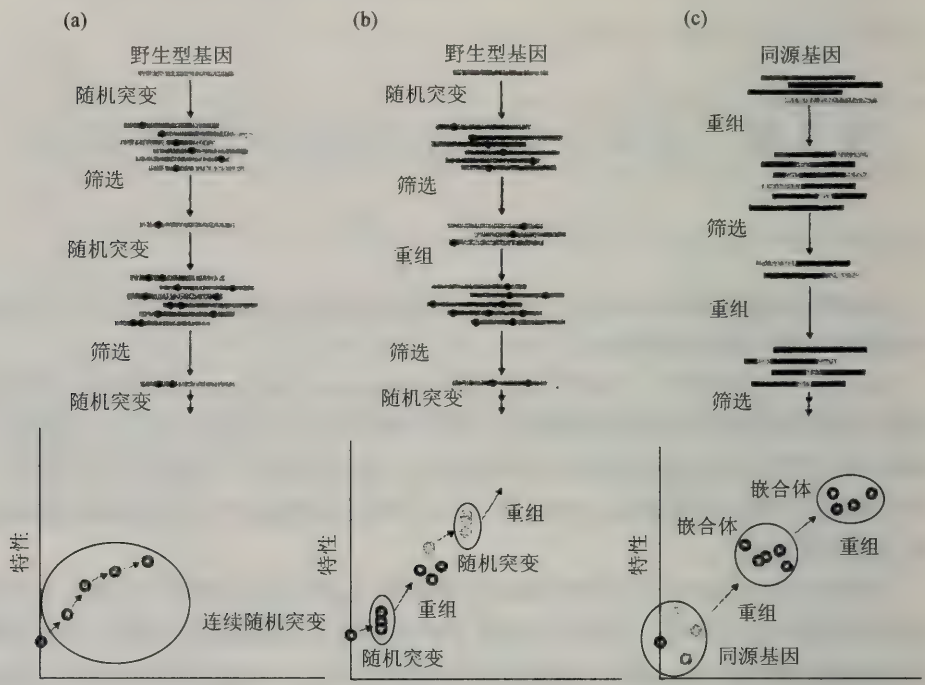


图 2.1　定向分子进化的基本策略[1]

(a)连续随机突变继之筛选出最佳改进的单一突变体,使每个循环含 1～2 有益突变的积累;(b)随机突变继之筛选两个或更多的改进的突变体重组,使有益突变组合并消除有害突变;(c)同源基因顺序重组(族改组)产生功能嵌合蛋白质,并且允许一步完成许多有益突变的组合,包括非累加突变

2.3　基 本 方 法

2.3.1　易错 PCR

易错 PCR(error-prone PCR, epPCR)是指在利用 Taq 酶进行 PCR 扩增目的基因的同时引入碱基错配,导致目的基因随机突变[2]。然而,经一次突变的基因很难获得满意的结果,由此开发出连续易错 PCR(sequential error-prone PCR)方法。即将一次 PCR 扩增得到的有益突变基因作为下一次 PCR 扩增的模板,连续反复地进行随机突变,使每一次获得的小突变累积而产生重要的有益突变。

在该方法中,遗传变化只发生在单一分子内部,故属于无性进化(asexual evolution)。它较为费力、耗时,一般适用于较小的基因片段(< 800bp)。此外,使用该方法易出现同型碱基转换。

2.3.2　DNA改组和族改组

DNA改组(DNA shuffling)又称有性PCR(sexual PCR),是蛋白质体外加速定向进化方法[3,4]。它分两步改进蛋白质的基因:①单一基因的突变,选择所要求的突变基因;②突变基因用DNase I片段化,继之用PCR体外重组(图2.2)。当DNA改

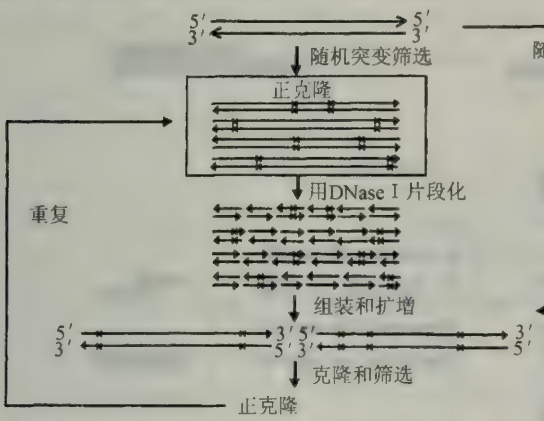


图2.2　DNA改组

它由两步重复循环组成:随机突变和突变的DNA改组。第一步(随机突变/筛选),由靶基因分离出几个突变体。第二步,突变基因用DNase I片段化,然后用没有引物的PCR组装,并用有引物的PCR扩增。用DNA改组产生DNA转化受体(一般是 *E. coli*),并选择出所要求表型的转化株

组用于一组同源基因时,这种技术称为"族改组"(family shuffling)(图2.3),族改组的重组子的产率是低的。当改组有80%核苷酸顺序同源的两个母体基因时,杂合体形成的频率少于1%,低的原因可能是由于异质复体形成的频率低[图2.4(a)]。为此,设计了两个改进的族DNA改组技术:①单股(ssDNA)的族改组。在这个方法中,制备两个同源基因的ssDNA,其中一个是一个基因的编码股,另一个是另一基因的非编码股。两个ssDNA分子用DNase I消化,它们的片段用于族改组。在第一轮杂化中,防止同源复体形成[图2.4(b)];②限制酶消化DNA片段取代随机片段DNA的改组。内切酶消化的DNA片段退火可以高频率地产生同源复体,

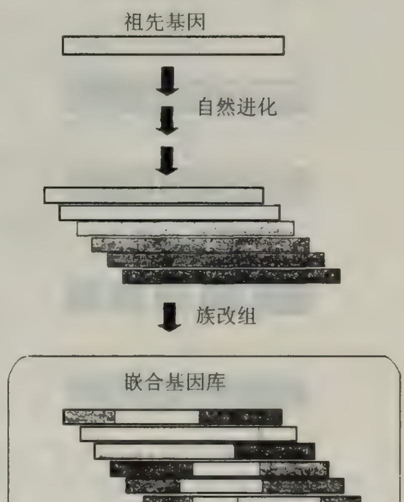


图2.3　族改组示意图

属于同一基因族的基因顺序改组产生嵌合基因库,并由库中选出所要求的嵌合体

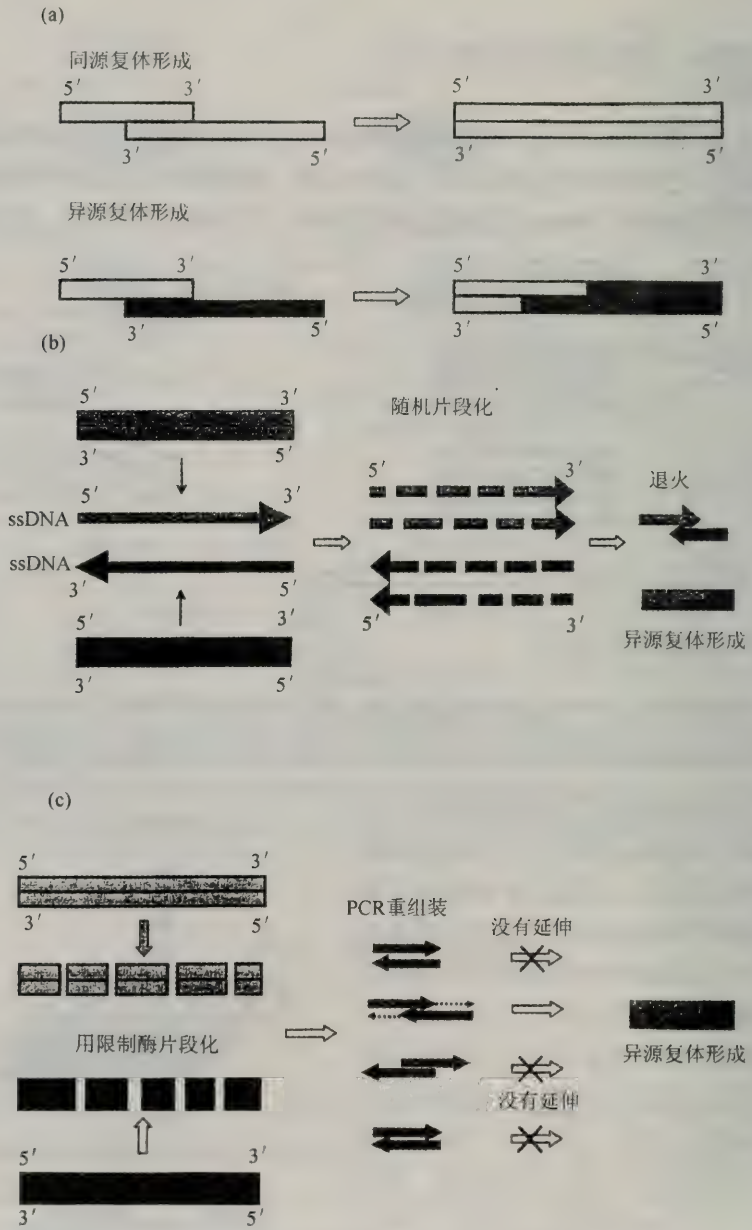


图 2.4 族改组中 DNA 杂合和延伸

(a) 两族基因的 DNA 股以空和实盒表示。在应用两个基因的族改组中，发生两种类型的退火：同源复体(来自同一基因的股退火)和非同源复体(来自两种不同基因的股退火)。如果同源复体形成的频率高于非同源复体，初始基因顺序再生往往优先出现，而不是嵌合基因的形成。(b) 为了防止在第一轮 PCR 中形成同源复体，ssDNA 由两个基因制备。这些 ssDNA 用 DNase I 片段化，只形成非同源复体分子。(c) 用限制酶分裂 DNA 片段的族改组中，形成同源复体和非同源复体，前者不出现 DNA 延伸，而后者发生 DNA 延伸，形成嵌合 DNA 片段

但是 DNA 延伸只发生在异质复体分子中[图 2.4(c),图 2.5]。

目前,族改组是各种来源的同源基因重组广为应用的技术,它是由 Stemmer DNA 改组法派生。族改组已用于多种同源基因产生功能嵌合蛋白质库[5]。例如,20 个以上分离的人干扰素-α-基因的改组产生嵌合库,它含有抗病毒和抗增殖活性都提高的各种突变体[6]。最好的嵌合体是来自 6 个母体基因顺序,其较之最好的

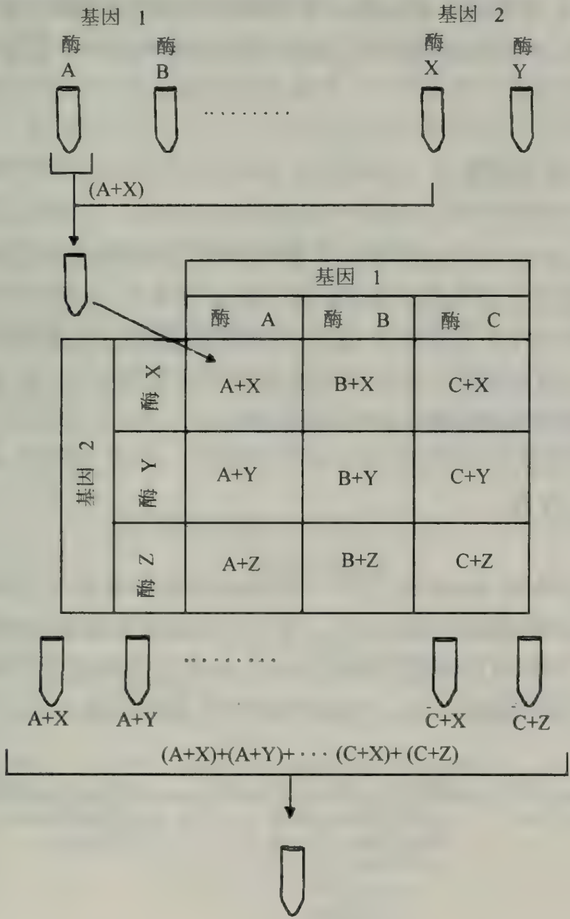


图 2.5　用限制酶分裂 DNA 片段的族改组

两个基因(基因 1 和基因 2)用几种限制酶分裂(对基因 1 是 A 和 B;对基因 2 是 X 和 Y)。两个限制的 DNA 样品,基因 1 的一个和基因 2 的另一个混合,进行无引物的 PCR 反应。限制的 DNA 样品重组见本图中间的表。在该步 PCR 之后,所有 PCR 混合物(A + X,A + Y 等)一起混合,并进行第二轮 PCR

母体抗病毒活性提高 87 倍。连续几轮 DNA 改组产生的嵌合体活性超过最好的人干扰素母体 185 倍。另一个例子是,26 种枯草杆菌蛋白酶的基因顺序(DNA 同源性仅为 56.4%,而基因改组要求 > 70%)被改组并对四种工业上重要酶学性质筛选[6]。尽管库中仅有 20%嵌合体是有活性的,但有的嵌合体的酶学性质优于最好的母体。此外,子代嵌合体呈现各种性质的组合。例如热稳定性和低温活性,而任何母体都不具有上述性质。最有趣的是,对某些进化性质而言,最好的母体与最好的嵌合体在顺序上不是紧密相关的。这表明最好的母体不一定是单一基因进化最好的起点。

2.3.3 外显子改组

外显子改组(exon shuffling)类似于 DNA 改组,两者都是在各自含突变的片段之间进行交换,前者尤其适用于真核生物。在自然界中,不同分子的内含子间发生同源重组,导致不同外显子的组合,是产生新蛋白质的有效途径之一。与 DNA 改组不同,外显子改组是靠同一种分子间内含子的同源性带动,而 DNA 改组不受任何限制,发生在整个基因片段上(详见第 6 章)。

2.3.4 合成改组

合成改组(synthetic shuffling)[7]是一种定向进化的新方法,母体的每个氨基酸可以同每个其他的氨基酸单独地组合。合成改组提供由数据库顺序到功能库的一条直接路线。15 种枯草杆菌蛋白酶基因的合成改组获得组合性质的嵌合酶,用其他定向进化方法得不到此结果。合成改组的设计策略见图 2.6。

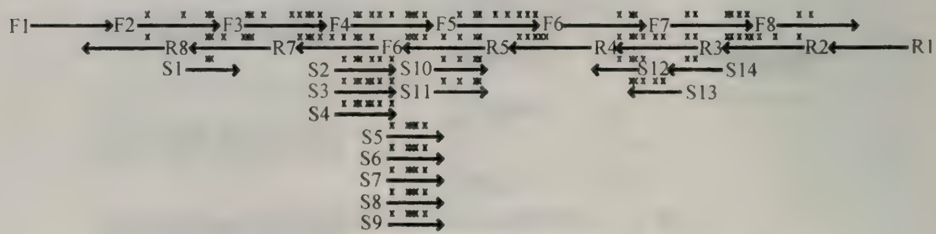


图 2.6 枯草杆菌蛋白酶合成改组的寡核苷酸设计策略

来自 15 种枯草杆菌蛋白酶母体的多样性被编码在全部 30 个寡核苷酸中。图中表示出 16 种骨架寡核苷酸(F1 ~ F8 和 R1 ~ R8)和 14 种"钉样"寡核苷酸(S1 ~ S14)相对位置。箭头指示 5′→3′方向,寡核苷酸位置用 X 表示

片段改组的突变体都与一个或更多的母顺序相关。相反,合成改组的突变体倾向于彼此远离,不再簇集在母体周围。合成改组使得突变体多样性增加,是由于破坏母体基因中所表达的氨基酸之间的键合,并改变氨基酸来自母体的倾向性。许多合成改组产生的蛋白质20%是有活性的,并具有新的功能。

合成改组使现有蛋白质的顺序空间扩大。合成改组在蛋白质库的构建上有很大的灵活性。由于所有的基因是合成的,密码子剂量可以被改变以便在宿主中表达,也可以消除母体基因中不同密码子的偏爱。对于低顺序同源性的母体基因,交换体也可以在单一氨基酸水平上产生。此外,合成的多样性不受母体基因的限制,这样,提供了由顺序数据库到蛋白质库的直接路线。在任何已知蛋白质中不存在的变化也可以工程到库中。因此合成改组是产生库的顺序多样性和功能多样性最有效的方法,很少受母体基因之间顺序同源性的限制。

2.3.5 体外随机引发重组[8]

体外随机引发重组(random-priming *in vitro* recombination,RPR)原理见图2.7。它是以单链DNA为模板,配合一套随机序列引物,先产生大量互补于模板不同位

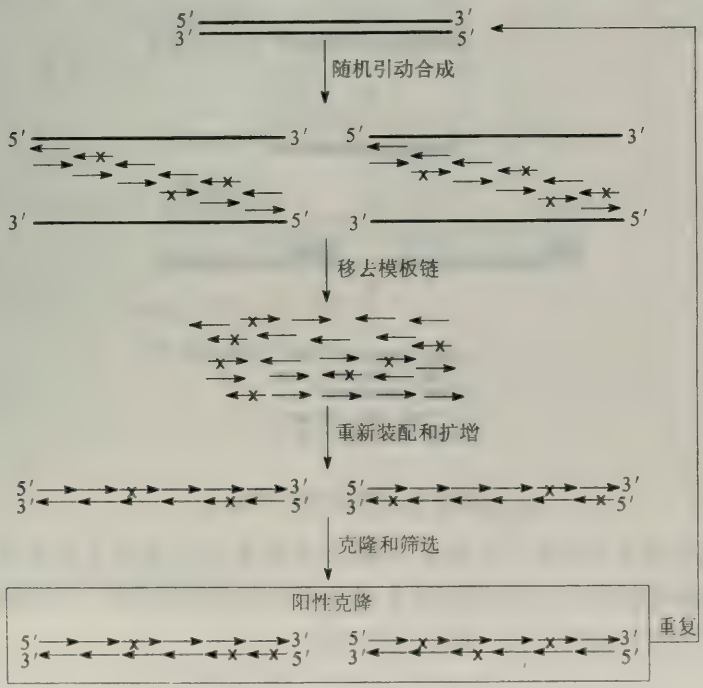


图2.7 体外随机引发重组原理图

点的短 DNA 片段,由于碱基的错配和错误引发,这些短 DNA 片段中也会有少量的点突变,在随后的 PCR 反应中,它们互为引物进行合成,伴随组合,再组装成完整的基因长度。如果需要,可反复进行上述过程,直到获得满意的进化酶。该法优于 DNA 改组法,其特点是:①RPR 可以利用单链 DNA 为模板,故可 10~20 倍地降低亲本 DNA 量;②在 DNA 改组中,片段重新组装前必须彻底除去 DNase I,RPR 则不需此步,故 RPR 方法更简单;③合成的随机引物具有同样长度,无顺序倾向性。在理论上,PCR 扩增使模板上每个碱基都应被复制或以相似的频率发生突变;④随机引发的 DNA 合成不受 DNA 模板长度的限制。

2.3.6 交错延伸[9]

交错延伸(stagger extension process, StEP)原理见图 2.8。它是在 PCR 反应中把常规的退火和延伸合并为一步,并大大缩短其反应时间(55℃,5s),从而只能合成出非常短的新生链,经变性的新生链再作为引物与体系内同时存在的不同模板退火而继续延伸。此过程反复进行,直到产生完整的基因长度,结果产生间隔的含不同模板序列的新生 DNA 分子。

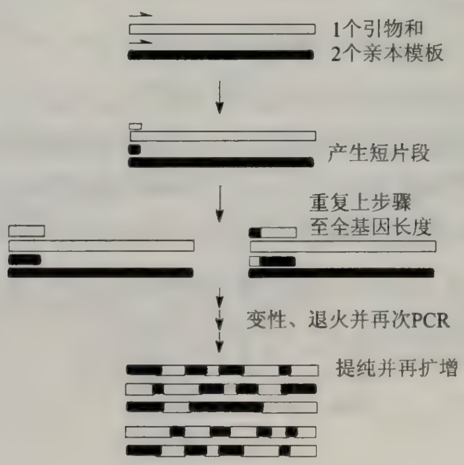


图 2.8　交错延伸原理图

StEP 法重组发生在单一试管中,不需分离亲本 DNA 和产生的重组 DNA。它采用的是变换模板机制,这正是逆转录病毒所采用的进化过程。该法简便且有效,为酶定向分子进化提供了又一强有力的工具。

2.3.7　酶法体外随机-定位诱变[10～12]

为解决空间结构未知酶的蛋白质工程问题，我们曾以类胰岛素样人参多肽基因和天冬氨酸酶基因为模型，探索了一种酶体外突变的新方法，即酶法体外随机-定位诱变（random-site-directed mutagenesis）（图 2.9）。该方法的内涵是对目的基因既采用随机突变增加突变位点，以快速产生优质酶，又让其突变受到一定限制，以减少筛选突变体的工作量。实际上，该法是通过控制 DNA 合成的底物种类和浓度比例实现碱基对的错配。该法又被称作 REM 法（random and extensive mutagenesis）。

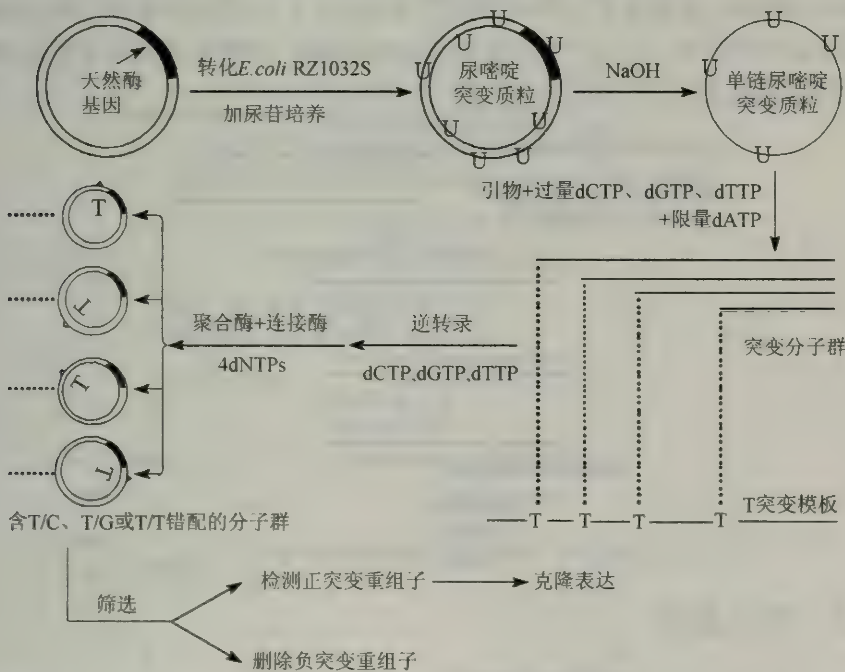


图 2.9　REM 法基本原理图

2.4　非同源基因改组

2.4.1　过渡模板随机嵌合

尽管族改组对于探索复杂定向进化问题是一种高效的策略，但仍有局限性。主要的限制是用 PCR 组装同源基因短的随机片段，要求同源性 > 70%，重组的顺

序之间连续顺序 10～15bp 长，仅有高度同源顺序可以重组，每个嵌合基因突变体仅有 1～4 交换体。此外，某些基因顺序尽管高度同源，有时难于重组。为了改进基因改组实验的结果，发展了一些非同源基因改组的方法[13,14]。

过渡模板随机嵌合(random chimeragenesis on transient template, RACHITT)[14～17]，它与 DNA 改组概念上明显不同。它不包括热循环、链转移或交错延伸反应，而是将随机切割的基因片段杂交到一个临时 DNA 模板上进行排序、修剪、空隙填补和连接(图 2.10)。其中的悬垂切割步骤使短片段(比 DNase 消化片段还短)得以重组，明显提高了重组频率；如果在片段重组前后采用易错 PCR 还可引入额外点突变。Coco 等[13]首次报道此法改造二苯并噻吩单加氧酶，产生的嵌合库平均每个基因含 14 个交换体，重组水平比 DNA 改组类方法(1～4 个交换体)高出几倍；并且可在短至 5bp 的序列同一区内产生交换体。这种高频率、高密度的交叉水平是 DNA 改组所难以达到的。

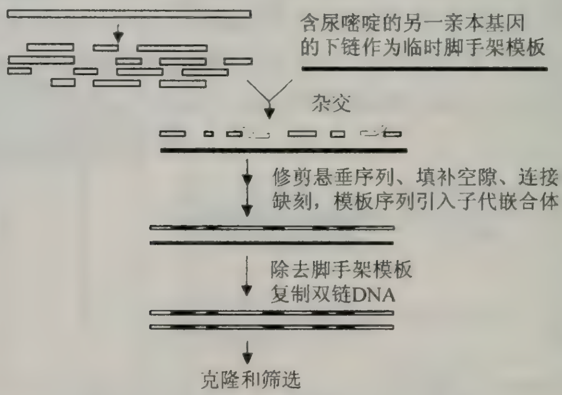


图 2.10 RACHITT 法基本程序示意图

2.4.2 增长截短

尽管过去几年发展了一些定向进化的新方法，但都局限于两种突变事件：碱基取代和同源重组。增长截短是通过缓慢的、定向控制 DNA 的消化，这样可以产生所有可能单一碱基对缺失的 DNA 片段库。

Luts 等[18]报道了构建杂合酶的“增长截短法”(incremental truncation for the creation of hybrid enzyme, ITCHY)。它是根据在两个目标基因片段之间产生单一碱基对缺失库的思想，继之这两个片段库随机融合(图 2.11)。该法作为模型系统已成功地用于 *E. coli* 甘氨酰胺核苷酸转移酶(PurN)和人(hGART)杂合体。作者应用 ITCHY 作为起点，导入 α-硫代核苷酸类似物(α-S-dNTP)以产生增长截短库(THIO-ITCHY)。这个方法是根据目标基因 PCR 扩增 α-S-dNTP 随机低频植入，然后外切

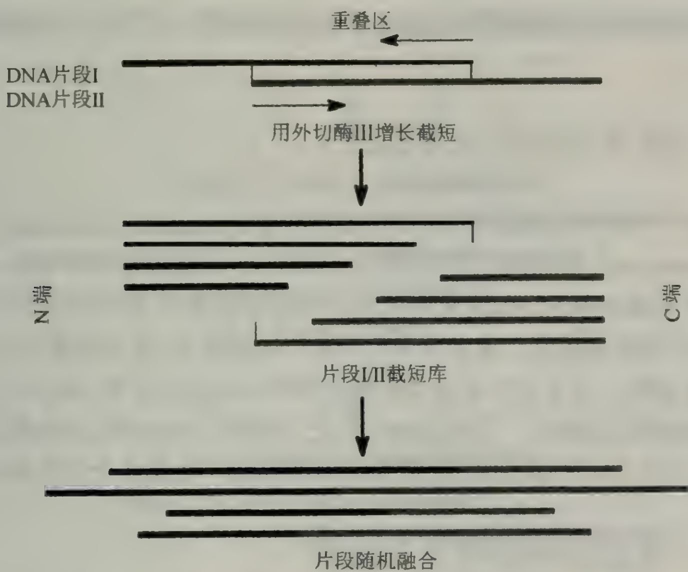


图 2.11　在两个 DNA 片段之间产生增长截短库一般图解

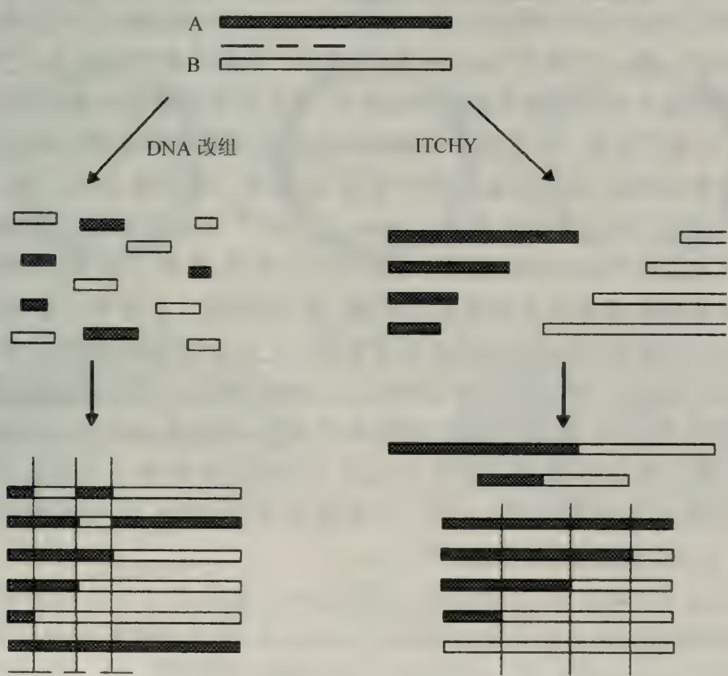


图 2.12　DNA 改组和 ITCHY 的比较

在 N 端有同源顺序的两个基因顺序(以黑线示之)将形成各种杂合库。DNA 改组产生多个交换体。组装库的长度分布很类似母体。N 端(基因 A)和 C 端(基因 B)的增长截短，继之片段库的连接，结果每个顺序形成单一交换体。如果 ITCHY 是在完整的基因上进行,融合库的长度则从零到 A + B

酶处理，结果在随机位点水解终止，并一步产生截短库。ITCHY 与 DNA 改组的比较见图 2.12。

2.4.3 非顺序同源蛋白质重组

Sieber 等[19]提出了非依赖同源顺序蛋白质重组(sequence homology-independent protein recombination, SHIPREC)(见图 2.13)哺乳动物细胞色素 P450 1A2(C 端部分)和可溶性细菌的 P450 BM3(N 端部分)，氨基酸顺序同源性仅约为 16%。通过两种基因随机截短形成的二聚体分子内环化产生嵌合库，继之在链节内(有限制位点)线性化并克隆。为了选择正确可读框可溶性的表达嵌合体，框内的片段同氯霉素乙酰基转移酶基因融合。应用 SHIPREC 已得到两个嵌合体，并表现出哺乳动物 P450 1A2 催化活性。作者指出，重复 SHIPREC 可以产生含多个交换体的嵌合库。

2.4.4 合理设计与随机组合方法

不依赖同源性的杂合体(包括整个结构域和亚结构域的融合)的合理设计有两个突出的例子。第一个例子，Hopfner 等[20]报道了胰蛋白酶和因子 Xa 之间的结构转换。两者都是由两个同源 β 桶结构域构成，杂合酶是将胰蛋白酶的 N 端融合到因子 Xa 的 C 端产生的。考虑两个母体酶的低蛋白质顺序同源性(38%)和由两者贡献的活性部位残基，杂合酶催化活性变宽，底物专一性明显不同。第二个例子，它可以认为是蛋白质工程的里程碑。Fersht 及同事[21]综合合理设计和 DNA 改组，应用吲哚甘油磷酸合酶(IGPS)的 β/α 桶结构作为骨架，作者工程 IGPS 的活性部位表现出磷酸核糖酰基邻氨基苯甲酸异构酶(PRAI)活性。上述两个酶催化色氨酸生物合成的连续步骤，并且在结构上是相关的。它们低的 DNA 顺序同源性(25%)不宜进行 DNA 改组。在初始合理设计时，在 IGPS 骨架上一系列缺失、取代和点突变产生可溶的 PRAI 产物并对 IGPS 底物有亲和力，但无催化活性。在第二相中，通过 DNA 改组和体外选择(应用 *E. coli* 的 PRAI 缺陷株)，其中一个杂合体(ivePRAI)的活性高于野生型 6 倍。尽管同原始的 IGPS 有 90% 顺序同源性，ivePRAI 完全丧失它的 IGPS 活性(图 2.14)。

通过综合不依赖同源(合理设计)和同源 DNA 改组，作者令人信服地证实了新设计策略的可行性，蛋白质骨架的初始的粗糙设计，继之用随机诱变和 DNA 改组精细调整可以成功地产生杂合酶，而不依赖母体顺序同源性。

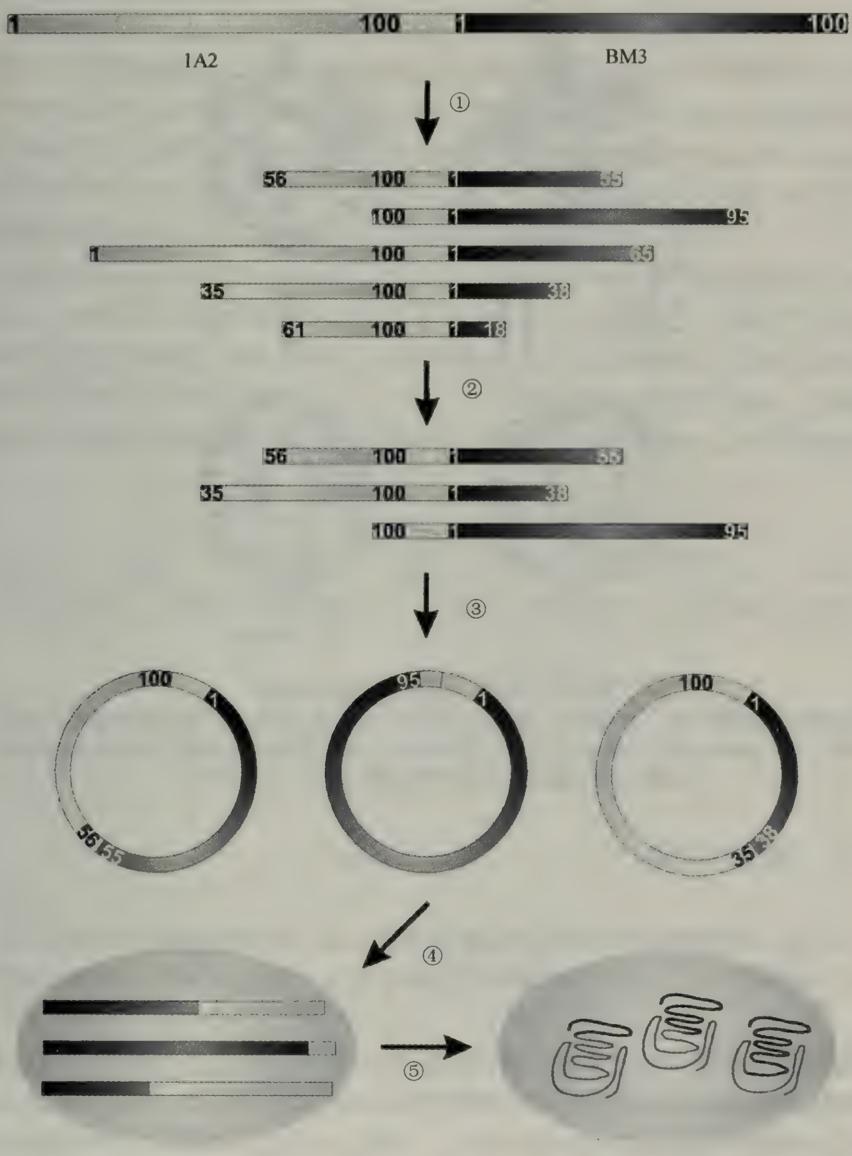


图 2.13　SHIPREC 操作示意图

基因二聚体是由蛋白质 1 的基因(1A2)(5′→3′)，链节顺序(含有效的限制位点)和蛋白质 2 的基因(BM3)组成。①二聚体片段化(例如，在 Mn^{2+} 存在下用 DNase I 消化)和处理以产生平端(例如用 S1 核酸酶或 T4 聚合酶)；②由库中分离单一基因长度的片段(例如用琼脂糖凝胶电泳)；③单一基因长度片段通过分子内平端连接环化；④通过链节区限制消化使环状 DNA 线性化，产生具有蛋白质 2 的 N 端部分和蛋白质 1 的 C 端部分嵌合基因库；⑤嵌合基因直接克隆到表达载体或者来自两个母体中一个末端引物进行PCR扩增后克隆到表达载体。图中数字代表基因长度的百分数。关键步骤是两个母体基因融合形成基因二聚体

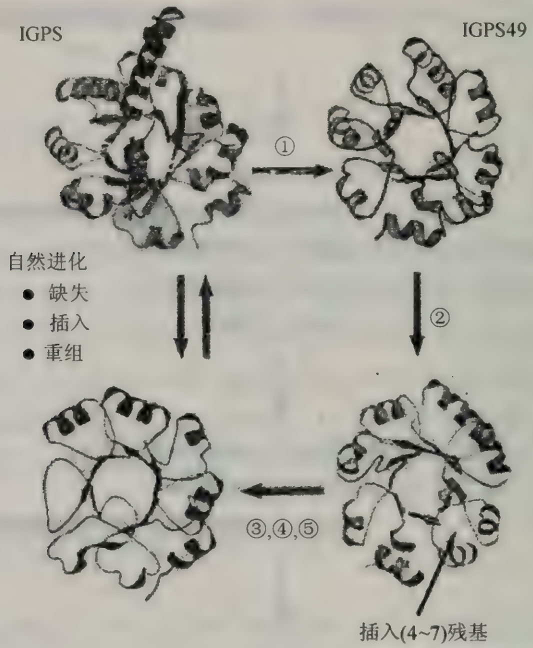


图 2.14 IGPS 骨架进化的实验策略
①IGPS 的 N 端缺失 48 个氨基酸产生 IGPS49;②相应于突环 β1α1 缺失 15 个氨基酸并有 4~7 残基取代,产生 IGPS49L1 库;③Asp 引入 184 位,并用 PRAI 一致顺序 GXGGXGQ 取代突环 β6α6;④体外选择;⑤体外重组(DNA 改组和 StEP PCR),产生 ivePRAI

2.4.5 模块组装

In silico 进化研究表明自然界的蛋白质进化较之同源重组进化的步伐大。在 Mante Carlo 模拟中,片段和结构域的非同源重组是产生全新功能蛋白质的成功途径。

过去几年更多的努力是进行多结构域酶的模块设计。锌指结构与内切核酸酶、甲基化酶和效应子结构域(转录活化剂/抑制剂)引起科学家更多的关注,这是由于它们可能对基因组研究有重要影响。锌指蛋白质是一类带有同锌连接的各种 Cys/His 基序(motif)的 DNA 结合蛋白质。在金属离子的结合上,锌指结构的结构域结合 DNA 的大沟。结晶学资料表明锌指同碱基三联体相互作用,锌指对 DNA 识别顺序可能通过几个关键氨基酸残基的改变而变化。

Barbas 及其同事[22]应用噬菌体展示技术鉴定整套 5′-GNN-3′三联体的特殊的锌指结构域。它们融合多个锌指产生专一指向复杂基因组中惟一的 DNA 顺序的结合模块,即 DNA 识别模块。3~6 个锌指结构域可识别 9~18 伸展的核苷酸顺

序，在低 nmol/L 范围表现出的亲和力，比单一碱基大 100 倍以上。

Chandrasegaran 和 Smith[23]应用锌指模块设计了带有非天然顺序特异性 IIS 型限制性内切核酸酶(RE)。天然限制酶的识别顺序通常限于 6 个核苷酸，由于基因组中多个识别位点存在而产生基因组 DNA 多点分裂。根据 IISRE 的模块性质，作者将 Fok1 的 F_N-内切核酸酶结构域融合到锌指模块上，获得的杂合体△QNK-F_N 有很高的活性，并在接近 DNA 识别位点的多个位置分裂。多个分裂产物可能不是识别顺序非专一性所致，而是锌指内切核酸酶缺乏结构刚性的结果。

聚酮合酶(polyketide synthase, PKS)和非核糖体肽合酶(nonribosomal peptide synthase, NRPS)的杂合体的产生代表了定向分子进化的另一领域[24]。PKS 和 NRPS 是由它们的模块结构来表征，由多个模块(它们本身是由不同催化活性的结构域组成)线型重排组装而成[图 2.15(a)]。在 PKS 和 NRPS 族中，缺失、添加或取代这些个别模块和结构域可以机械地设计上述分子。Ranganathan 等[25]报道定向

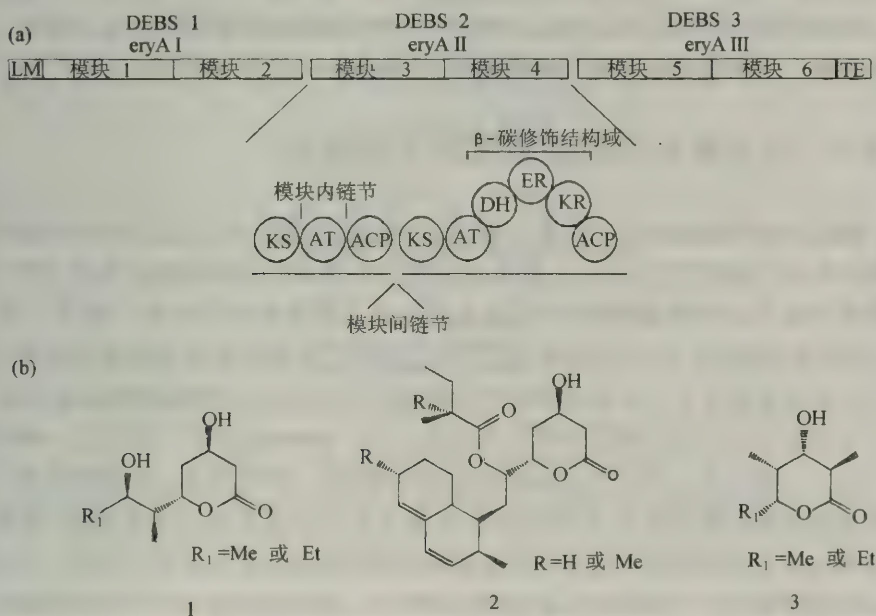


图 2.15　PKS 多模块线型重排

(a)DEBS 的结构。酶复合物被分为 3 个亚基(eryA I ~ III)，每个由两个模块组成。此外，第一个亚基(eryA I)携带一个负载的结构域(LM)，而 eryIII 有一个释放的结构域(TE)。亚基相互作用是非共价蛋白质-蛋白质相互作用，亚基内的模块是共价连接(模块间链节)。对功能而言，每个模块最少需要 3 个结构域：一个酮基合成酶(KS)，一个酰基转移酶(AT)和酰基载体蛋白(ACP)。此外，β-碳修饰酶，例如酮基还原酶(KR)，脱氢酶(DH)和烯酰基还原酶(ER)。(b)通过杂合体 PKS 合成目标结构的例子。结构 1 是一个抑制素族(2)的生物活性类似物(一类降胆固醇化合物)。结构 3 代表 eryA 模块的天然产物，其是杂合体复合物设计的起点。Et——乙基；Me——甲基

设计杂合体 PKS。结构 1[图 2.15(b)]是降低胆固醇的抑制素(statin)族(2)的一种类似物。从 deoxyerythronotide B synthase (DEBS)出发,纳巴霉素(模块 11/12)的各种取代的结构域和模块导入以改变原始的 eryA I triketide 产物和结构 3 到 1。最初试图用 rapPKS 模块 12 取代 eryA I 的完整模块 2,通过模块间链接节区的融合,结果未检出 triketide。根据假说,模块间链节区不利于产生功能杂合体蛋白质,作者用 rapPKS 的个别结构域取代了 eryA I 模块,产生杂合体 PKS(产物 I),但产量低。Tang 等[26]报道了在 3 个相关的 PKS 工程中用模块间链节出现了相似问题。Gokhal 等[27]成功地通过保留工程分析和重组个别的 eryA 和 rif 模块,在模块间链节区保存母体顺序,结果形成的杂合体蛋白质有完全的活性。

对于特殊化合物的生产,除了杂合体复合物设计外,PKS 和 NRPS 复合物的模块性质也为组合途径提供了可能性。McDaniel 等[28]通过 DEBS 中单一或多个酰基转移酶(AT)、酮基还原酶(KR)、脱氢酶(DH)和烯酰基还原酶(ER)结构域取代产生 60 多种红霉素衍生物库。应用 DEBS 作为模型系统,作者把 3 个亚基(eryA I ~ III)克隆到质粒中,并在每个质粒中交换 AT 和 β-碳修饰结构域。经一系列的转化,库的多样性倍增而不是加和。PKS 和 NRPS 族的结构域和模块组合成单一复合物将进一步扩展酶合成复合物的能力,但这只是代表该方向的第一步[29,30]。

2.4.6 任意数量的碱基随机插入和缺失

随机突变同高通量筛选的组合是改进蛋白质功能或产生人工酶的普通策略。在体内外引入随机突变方法中,易错 PCR 突变是普遍采用的随机点突变方法。然而易错 PCR 方法有氨基酸倾向性,因为只取代三联体密码子的单一碱基。例如,由 AUG(蛋氨酸)到 UAG(色氨酸)是不可发生的。寡核苷酸指导的诱变和盒式诱变可以在氨基酸水平上实现非倾向性随机取代,然而它们不能在随机位置上突变。一些合成方法也只适于靶基因窄范围内的突变。Murakami 等[31]提出了一种有效的方法——随机插入/缺失(random insertion/deletion, RID)突变。在随机位置上沿着靶基因的整个范围突变,能够在随机位置上缺失任意数量的连续碱基(直到 16 个碱基),并且同时把任意数量碱基的特异顺序或随机顺序插入同一位置。用该法可以随机地取代 3 个选择的碱基,特异的密码子、混合密码子或非天然氨基酸的 4 个碱基密码子。RID 突变的关键步骤示如图 2.16。

至今,普遍用易错 PCR 引入突变,继之 DNA 改组和其他次级操作。在 PCR 方法中,只有单一碱基突变,这限制了移入氨基酸的多样性,单一突变往往不能导致希望的氨基酸取代。在这方面,通过 RID 突变所产生的库同常规技术比较有更大的潜力。RID 突变方法可以用于(例如丙氨酸扫描)研究蛋白质折叠或酶活性的关键位置。同样,RID 突变可以把单一随机氨基酸引入随机位置,这是常规方法难以

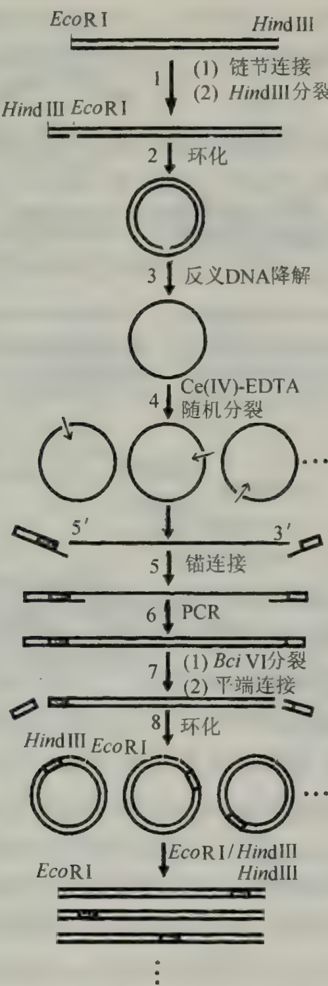


图 2.16　随机插入/缺失(RID)突变一般图解

它有 8 个主要操作步骤:第一步:用 *Eco*RI 和 *Hin*dIII 消化出发基因所得片段与链节连接;产物用 *Hin*d III 消化制成反义链有缺口的线性 dsDNA。第二步:基因片段用 T4DNA 连接酶环化制成反义链有缺口的环状 dsDNA。第三步:环状 dsDNA 用 T4DNA 聚合酶处理产生环状 ssDNA。第四步:环状 ssDNA 用 Ce(IV)-EDTA 复合物处理,随机在单一位置分裂。第五步:线性 ssDNA 两端带有未知顺序,分别连接到 5′锚位和 3′锚位。第六步:与两端的两个锚位连接的 DNA 用 PCR 扩增。第七步:PCR 产物用 *Bci*VI 处理,由 5′锚位(在 5′端)去除 7 个碱基。*Bci*VI 处理也删去 3′端的一定数量的碱基。第八步:消化的产物用 Klenow 片段处理制成平端,并且用 T4DNA 连接酶再次环化。产物用 *Eco*RI 和 *Hin*dIII 处理,并且所得片段克隆到修饰的 pUC18(pUM)的 *Eco*RI ~ *Hin*dIII 位点

办到的。

RID 突变也可以把任意数量碱基引入靶基因。例如限制位点(AGATCT)可以

引入 3 个连续的碱基。基本上不限制引入目标基因的碱基数量。当用限制酶 *Bci*VI 时,碱基数可以由 3′端缺失 0 ~ 6,如果用其他的限制酶,例如 *Bsg* I,缺失范围将延伸到 16。

RID 突变另一个重要的应用是随机移入非天然氨基酸,其可以由四碱基密码子编码[32,33],即 GGGU、CGGU、CUCU、CCCU 和 AGGU,它们可以分别指定为不同的非天然氨基酸。由于引入非天然氨基酸大的侧链基团,非天然突变的一个主要问题是突变体往往错误折叠。非天然突变体的 RID 库的构建和突变体的选择将产生有益功能的蛋白质。

总之,RID 突变发展了突变体库。RID 突变也能够在目标基因随机位置上插入任何数量的碱基或缺失直到 16 个碱基。该方法将会广泛地用于生命科学和生物技术。

2.4.7 片段突变

片段突变是一个重要的随机突变方法。应用 PCR 可以产生基因缺失或插入。Pikkemaat 等[34]提出了体外缺失和重复的完全随机的方法。通过组合两个独立产生的一批 DNA,含有 3′或 5′截短的基因片段,产生随机融合库,结果在融合点产生缺失和重复。通过控制启动截短的水平,可以控制片段突变,以限制库的大小(图 2.17)。该策略同随机突变和基因改组的组合可能对定向分子进化有导向作用。Pikkemaat 等用自养黄杆菌(*Xanthobacter autotrophicus*) GJ10(Dh1A)产生的卤代烷脱卤酶作为模型系统,当在 1,2-二氯代乙烷(DCE)上生长时,脱卤酶是细菌代谢途径中的关键酶。它可以将氯代和溴代化合物转为相应的醇,在环境保护方面有重要意义,因为其底物是环境污染物。在该酶顺序中的重复和缺失已知在适应新底物起重要作用。实验结果表明了该方法的有效性,也表明了该突变方法在定向分子进化实验上的潜在价值。

2.4.8 设计的寡核苷酸组装

ADO(assembly of designed oligonucleotide)法是根据顺序信息指导各种寡核苷酸设计,控制所设计的寡体的重叠,组装成高重组频率的全长产物[35]。ADO 法使母体基因自身杂化减少到最低,用于酶定向分子进化明显超过常规的族改组。ADO 已用于 *Bacillus subtilis* 两种脂肪酶族基因(*LipA* 和 *LipB*)改组。图 2.18 说明 ADO 的一般概念。根据顺序空间的同源性,应用两种策略进行组装:①两种基因 A 和 B 被匹配重组;图中各种颜色星号给出各种编码氨基酸的密码子。在母体基因的同样位置有两种颜色星号寡核苷酸片段表示具有简并性的合成寡核苷酸片

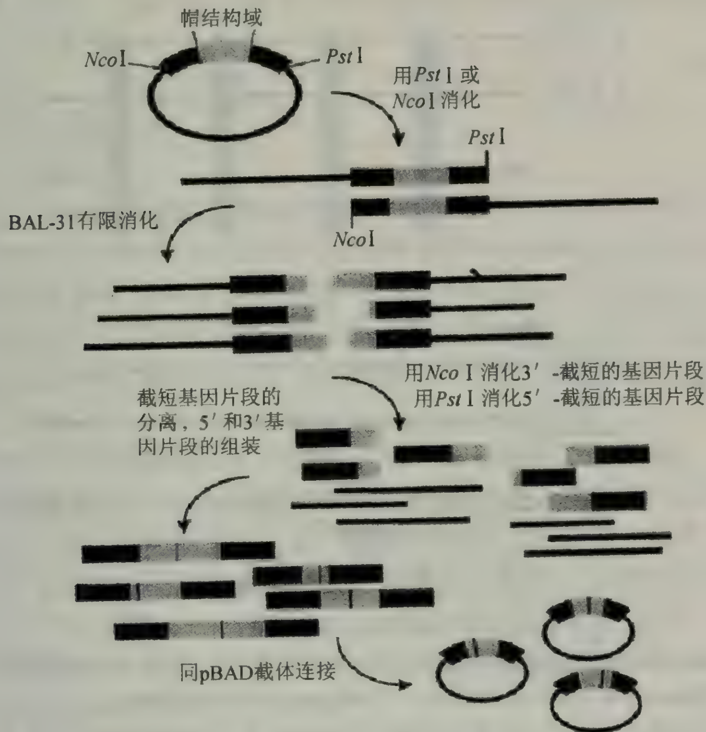


图 2.17　片段突变操作的图解说明

载体首先在脱卤酶基因的 5′或 3′端线性化，外切核酸酶作用逐渐去除剩余的载体 DNA，产生一批 3′或 5′截短的基因片段。在两端组装后，片段突变库被连接到新的载体上，并转化到 *E. coli*，对酶活性进行筛选

段。灰色部分表示顺序的保守区，其可以用作同源重组的链节；②在 oligose 两侧之间没有同源性。它可以在未知末端顺序 ssDNA 股之间连接组装。

寡核苷酸的设计是按以下方式完成的。根据 *LipA* 和 *LipB* 的蛋白质匹配，氨基酸顺序被分为 5 个单元。根据标准编码剂量，将氨基酸残基写成核苷酸，而以 *LipA* 和 *LipB* 的最初 DNA 顺序作为参考。寡核苷酸的 5 个单元 Lip1 ~ Lip5 按顺序设计。当两个顺序中的氨基酸不同时，应用简并核苷酸。例如，如果 Ser 是被一个密码子(TCT)在某一位置编码，在第二个母体基因产物 Lys 是由另一个密码子(AAA)在上述位置编码，那么这个位置的核苷酸应选择(T/A)(C/A)(T/A)。在这种情况下，在这个位置编码氨基酸的核苷酸将是 WMW(按 IUB 编码)。遗传密码的简并性可以把额外编码的氨基酸引入到既定位置。例如，TA(T/A)可以编码出 Tyr 或终止密码子，AC(T/A)可以编码出 Thr。这样，“额外”氨基酸提高了编码的多样性。

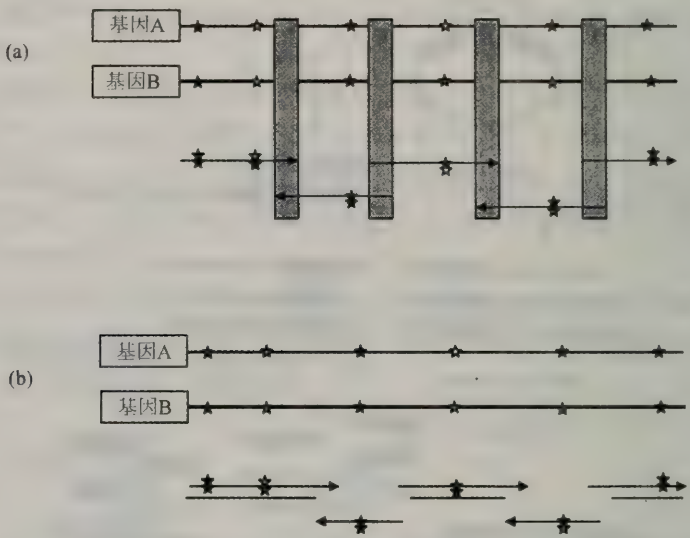


图 2.18 ADO 的一般概念

片段连接的两种可能策略:(a)两个基因 A 和 B 改组匹配。不同星号给出编码各种氨基酸的密码子,而在母体基因同一位置有两种星号的寡核苷酸片段表示具有简并性的寡核苷片段。空心星号表示顺序的保守区,它可以作为同源重组的链节。(b)在两侧寡核苷酸之间没有同源性,它可以在未知末端顺序 ssDNA 股之间连接组装

片段的 PCR 组装是以直接方式进行的,正常情况下分两步。在 PCR 反应中,合成的寡核苷酸片段 Lip1 ~ Lip5 代表单股设计合成的 oligose。两侧片段的方向相反,例如 Lip1 是 5′→3′,Lip2 是 3′→5′,Lip3 是 5′→3′等。加上 dNTP,缓冲液和聚合酶,在几轮扩增之后,各种片段彼此连接并形成双股 DNA。全基因的扩增示如图 2.19。全基因被连接到高表达载体 pET22b(+)并转化到 *E. coli* BL21(DE3)。

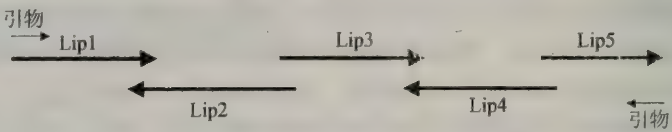


图 2.19 设计的寡核苷酸组装过程

Lip1→Lip5 代表设计的寡核苷酸,箭头表示片段的方向

应用 ADO 法产生 3000 脂肪酶突变库,从中筛选出内消旋双乙酸酯对映体选择性明显提高的突变体。

ADO 法的主要优点是母体基因的自身杂化最低或消除,提高库的质量,减少

筛选工作量。这样,ADO 可以预期是一种功能蛋白质和代谢工程定向进化的一种积极方法。

2.4.9 体外区室化

实验室定向进化是模拟自然进化的一种强有力的策略。自然和定向进化都需要基因型和表型之间的联系。在体外,这种联系通常是由基因与蛋白质来体现,并通过各种技术来完成,包括噬菌体、病毒、细菌和酵母展示,质粒展示,核糖体展示和 mRNA-肽融合。这些展示技术证明在结合蛋白质选择上是很成功的。相反,通过展示途径,选择酶的成功率则小。间接选择,例如通过与过渡态类似物或酶抑制剂结合,一般很难产生有效的催化剂。

Griffiths 等[36]根据体外区室化(IVC)发展了一种酶体外选择系统。各种反应被分为显微镜区室,每个仅约 5fl,形成油包水乳状液,50μl 反应液可以被分散到约 10^{10}含水区室,可以选择许多基因,这个系统是高敏感和经济的。作者以前用 IVC 选择 DNA 甲基转移酶[37]和 Taq DNA 聚合酶的进化酶[38]。作者根据产生微珠库提出一种新的 IVC 策略,每个微珠展示一个基因和它编码的蛋白质。“微珠展示库”可以根据展示的多肽的结合活性来选择,或在多重转换条件和选择反应环境,对可溶的非 DNA 底物的催化进行选择。

定向进化策略对于许多非合理设计来讲是成功的。现在研究人员集中在发展新筛选和非同源重组技术以及复杂系统。例如途径,病毒,甚至完整基因组的定向进化。由基因组计划所获得的大量新基因顺序将为新生物催化剂和代谢途径的定向进化提供选择基因的“宝库”。

不依赖于同源性方法的开发仍处于早期阶段。合理设计虽然取得了积极的结果,但它取决于广泛的结构和功能信息,因而限制它的广泛应用。不依赖同源性方法将增加组合蛋白质库的多样性,将产生许多有意义的和未预料到的结果,将最终导致成功地设计新酶以及更好地了解酶的结构与功能关系。

参 考 文 献

1 Schmidt-Dannert C. Directed evolution of single protein, metabolic pathways and viruses. Biochemistry, 2001, 40: 13125 ~ 13136

2 Chen K, Arnold F H. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of substilisin E for catalysis in dimerthylformamide. Proc Natl Acad Sci USA, 1993, 90: 5618 ~ 5622

3 Stemmer W P. DNA shuffling by random fragmentation and reassembly-*in vitro* recombination for molecular evolution. Proc Natl Acad Sci USA, 1994, 91: 10747 ~ 10751

4 Stemmer W P. Rapid evolution of a protein *in vitro* by DNA shuffling. Nature, 1994, 370: 389 ~ 391

5 段招军,赵军,王红艳等. α 型干扰素家族定向进化文库的构建及细胞筛选. 病毒学报,2001,17:172 ~ 174

6 Ness J, Welch M, Gilver L, *et al*. Creation of a functionally diverse enzyme library by DNA family shuffling. Nat Biotechnol, 1999, 17: 893 ~ 896

7 Ness J E, Kim S, Gottman A, *et al*. Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. Nat Biotechnol, 2002, 20: 1251 ~ 1255

8 Shao Z, Zhao H, Arnold F H. Random-priming *in vitro* recombination: an effective tool for directed evolution. Nucleic Acids Research, 1998, 26 : 681 ~ 683

9 Zhao H, Giver L, Shao Z, *et al*. Molecular evolution by staggered extension process (StEP) *in vitro* recombination. Nat Biotechnol, 1998, 16: 258 ~ 261

10 张红缨,李正强,张今. 酶法体外建立天冬氨酸酶基因的随机突变库. 科学通报,1991,36:1500 ~ 1502

11 Zhang H Y, Zhang J, Lin L. Enhancement of the stability and activity of aspartase by random and site-directed mutagenesis. Biochem Biophys Res Commm, 1993, 192: 15 ~ 19

12 Ling M M, Robinson B M. Approaches to DNA mutagenesis: an overview. Anal Biocnem, 1997, 254: 157 ~ 178

13 Coco W M, Levinson W E, Crist M J, *et al*. DNA shuffling method for generating highly recombined genes and evolved enzymes. Nat Biotechnol, 2001, 19: 354 ~ 359

14 徐卉芳,张先恩,张用梅. 体外分子定向进化研究进展. 生物化学与生物物理进展,2002,29:518 ~ 521

15 肖志壮,刘梦海,汪天虹,曲音波. 蛋白质定向进化的研究进展. 生物工程进展,2001,21:31 ~ 33

16 赵志虎,马清钧. DNA 重排及体外分子进化. 生物技术通讯,2000,11:275 ~ 280

17 胡军. 酶技术发展与酶定向进化. 工业微生物,1999,29:37 ~ 42

18 Lutz S, Ostermeier M, Benkovic S J. Rapid generation of incremental truncation libraries for protein engineering using alpha-phosphothioate nucleotides. Nucleic Acids Research, 2001, 29: e16

19 Sieber V, Martinez C A, Arnold F H. Libraries of hybrid proteins from distantly related sequence. Nat Biotechnol, 2001, 19: 456 ~ 460

20 Hopfner K P, Kopetzki E, Kresse G B, *et al*. New enzyme lineages by subdomain shuffling. Proc Natl Acad Sci USA, 1998, 95: 9813 ~ 9818

21 Altamirano M M, Blackburn J M, Aguayo C, *et al*. Directed evolution of new catalytic activity using the alpha/beta-barrel scaffold. Nature, 2000, 403: 617 ~ 622

22 Segal DJ, Dreier B, Beerli R R, *et al*. Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5′-GNN-3′ DNA target sequences. Proc Natl Acad Sci USA, 1999, 96: 2758 ~ 2763

23 Chandrasegaran S, Smith J. Chimeric restriction enzymes: what is next. Biol Chem Hoppe-Seyler, 1999, 380: 841 ~ 348

24 Mootz H D, Marahiel M R. Design and application of multimodular peptide synthetases. Curr Opin Biotechnol, 1999, 10: 341 ~ 348

25 Ranganathan A, Timoney M, Bycroft M, *et al*. Knowledge-based design of bimodular and trimodular poluyketide synthases based on domain and module swaps: a route to simple statin analogues. Chem Biol, 1999, 6: 731 ~ 741

26 Tang L, Fu H, McDaniel, R. Formation of functional heterologous complexes using subunits from the picromycin, erythromycin and oleandomycin polyketide synthases. Chem Biol, 2000, 7: 77 ~ 84

27 Gokhale R S , Tsuji S Y, Cane D E, *et al*. Dissecting and exploiting intermodular communication in polyketide synthases. Science, 1999, 284: 482 ~ 485

28 McDaniel R, Thamchaipenel A, Gustafsson G, *et al*. Multiple genetic modifications the erythromycin polyketiale synthase to produce a library of novel unnatural products. Proc Natl Acad Sci USA, 1999, 96: 1846 ~ 1851

29 Liou G F, Knosla C. Building-block selectivity of polyketide synthases. Current Opinion in Chemical Biology, 2003, 7: 279~284

30 Shen B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. Current Opinion in Chemical Biology, 2003, 7: 285~295

31 Murakami H, Hohsaka T, Sisido M. Random insertion and deletion of arbitrary number of bases for codon-based random mutation of DNAs. Nat Biotechnol, 2002, 20: 76~81

32 Hohsake T, Ashizuka Y, Sasaki H, *et al*. Incorporation of two different nonnatural amino acids independently into a single protein through extension of the genetic code. J Am Chem Soc, 1999, 121: 12194~12195

33 Hohsaka T, Kajihara D, Ashizuka Y, *et al*. Efficient incorporation of nonnatural amino acids with large aromatic groups into streptavidin in *in vitro* protein synthesizing systems. J Am Chem Soc, 1999, 121: 34~40

34 Pikkemaat M G, Jansen D B. Generating segmental mutation in haloalkane dehalogenase: a novel part in the directed evolution toolbox. Nucleic Acids Research, 2002, 30: 35

35 Zha D X, Eipper A, Reetg M T. Assembly of designed oligonucleotides as an efficient method for gene recombination, A new tool in directed evolution. Chem Biochem, 2003, 4: 34~39

36 Griffiths A D, Tawfik D S. Directed evolution of an extremely fast phosphotriesterase by *in vitro* compartmentalization, EMBO J, 2003, 22: 24~35

37 Lee Y F, Tawfik D S. Griffiths A D. Investigating the target recognition of DNA cytosine-5 methyltransferase *Hha*I by library selection using *in vitro* compartmentalization (IVC). Nucleic Acids Research, 2002, 30: 4937~4944

38 Ghadessy F J, Ong J L, Holliger P. Directed evolution of polymerase function by compartmentalized self-replication. Proc Natl Acad Sci USA, 2001, 98: 4552~4557

第3章　定向分子进化的选择/筛选策略和方法

3.1　引　　言

第2章介绍了定向分子进化的策略和方法，核心是讨论如何增加多样性，拓展顺序空间，积累有益突变。这仅仅是定向分子进化的一个重要方面。另一个重要方面是在产生多样性之后，即在产生各种突变体库之后；所希望的酶必须从库中分离出来，这可以通过选择或筛选操作实现。

选择(selection)和筛选(screening)往往容易混淆。选择是利用所希望的突变体的独有的生存条件，模拟达尔文进化过程，由大的组合库中鉴定突变体；筛选则是检验大的组合库中所有成员来鉴定突变体，是一种主动的搜索过程。选择/筛选是分子生物学常用的方法。图3.1表示应用正选择克隆载体的一种体内选择方法。它分为随机筛选，易化筛选(facilitate screening)和选择。

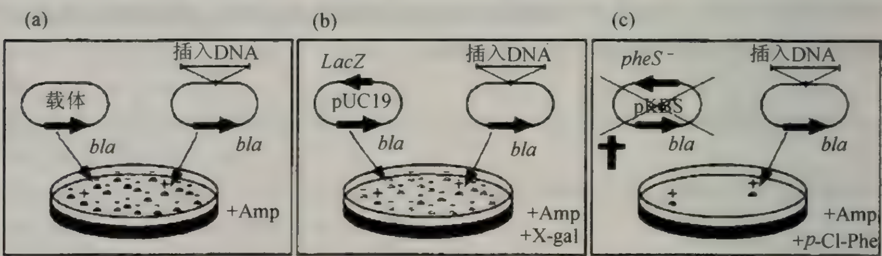


图3.1　变异克隆的搜索策略

转化后营养琼脂平板上"+"表示含有插入基因的质粒；"-"表示仅有载体DNA；*bla* 基因编码β-内酰胺酶，其抗氨苄青霉素(Amp)。(a)随机筛选。所有克隆的表型是相同的。分离随机挑选的克隆DNA，分析所希望的质粒。(b)易化筛选。克隆载体pUC19携带编码β-半乳糖苷酶的 *LacZ* 基因片段，编码多肽(酶的α-片段)可以同细胞的 *LacZ* 片段产物结合形成功能β-半乳糖苷酶，其分裂5-溴-4-氯-3-吲哚基β-D-吡喃半乳糖苷(X-gal)产生靛蓝衍生物沉淀。在含有X-gal琼脂板上呈蓝色。相反，希望得到的克隆是无色，这是由于克隆的基因插入到α-片段的编码区，因此使β-半乳糖苷酶活性丧失。由蓝色克隆背景挑选白色克隆(图中黑斑)以找出正确基因型。(c)选择。正选择载体pKBS携带突变体 *pheS* 基因，编码苯丙氨酰基-tRNA合成酶α-亚基，给出松弛的底物特异性。在苯丙氨酸类似物对-氯-苯丙氨酸(P-Cl-Phe)的存在下，由于类似物掺入蛋白质造成总的蛋白质功能失活，表达突变体pheS菌株不能生存。然而在营养琼脂板上转化的细胞可以形成含有Amp和P-Cl-Phe，如果条件致死pheS是由于克隆的DNA片段的插入，这样，在转化后所有生长的克隆都有正确的基因型

选择的优点是可以检验更大的库。一般来讲，可以进行选择克隆数超过先进的筛选方法5个数量级。可靠的选择动力是宿主的生存直接与要求的表型偶联。这也是对大多数选择方法的主要挑战。尽管所检验的克隆数低，但筛选操作日显重要。一个重要原因是方法的巨大进步，现在已达到自动化、小型化，筛选酶的各种工作站都已建立起来。

根据酶活性进行筛选(或选择)是最普遍采用的方法，或者在微量滴定板上直接检验颜色或荧光反应产物的形成，或者是固相检验(例如，琼脂板或滤板)，在琼脂板或滤板上克隆的固相筛选可以迅速地筛选大库，或根据存在于固相中底物的分裂或通过局部pH变化用pH指示剂检验产物的释放来鉴定各种突变体，通过改变固相中释放产物的溶解度所产生的亮区和沉淀的发展，或监控决定不溶底物分裂的生长速度等。目前大多数固相筛选方法缺乏催化活性的灵敏定量，因此适合于第一轮定向进化的筛选。迅速灵敏的荧光高通量数字呈像筛选已得到发展。如果适当的荧光或生色底物是有效的话，数字呈像是大库定量固相筛选强有力的方法。应用平板阅读器定量比色测定或荧光测定，直接检验全细胞或裂解液或表达的突变体酶的定位和底物可及性的培养上清液。虽然耗时，但可根据进化所要求的功能进行检验。在某种情况下，所要求的新催化活性可以直接与表达宿主的表型联系起来，例如宿主营养缺陷型或提供抗生素或细胞毒试剂，通过体内选择可以筛选大库($>10^6$)，表3.1列出筛选酶的某些方法。

酶定向分子进化选择/筛选的各种策略和方法似乎都是根据表型和活性或它们的偶联衍生的，它贯穿在本章所介绍各种策略和方法之中，因此本书不做专题讨论。

过去10年，展示技术已成为肽和蛋白质配体以及描述体内相互作用模型的主要工具[1~4]。噬菌体和核糖体展示系统主要用于发现，而酵母双杂交方法主要用于体内相互作用分析。

噬菌体展示和酵母双杂交系统需要体内步骤，因而应用受到一定的限制。在噬菌体展示中，库必须转化到细菌中，可能的独立顺序量限于 $10^9 \sim 10^{10}$。表达的顺序总量可能由于其他一些因素干扰进一步减少。这些因素包括未折叠分子的降解，在细菌宿主中低表达，缺乏噬菌体表面的加工，在 *E. coli* 氧化的围膜间隙缺乏折叠和基因产物的毒性等。同样，酵母双杂交系统需要把相互作用伴侣克隆到酵母中，所检验的量局限于 $10^6 \sim 10^7$。此外，双杂交途径，相互作用必须发生在核内，限制更严格。体外技术，例如核糖体和mRNA展示，克服了噬菌体展示和双杂交系统的许多限制。该途径偏向性低，由于不需要转化步骤，可能的独立顺序量 $>10^{12}$。

表 3.1 筛选酶的方法举例

酶	检验方法
蛋白酶	晕圈(A)
蛋白酶	吸收(S)
蛋白酶	FRET(S)
蛋白酶	FP
枯草杆菌蛋白酶	吸收(S)
蛋氨酸氨基肽酶 D-丙氨酰-D-丙氨酸二肽酶	吸收(S)
青霉素 G 酰化酶	颜色(A,S)
青霉素 G 酰化酶	荧光(S)
青霉素 G 酰化酶	荧光(S)
葡萄球菌脂肪酶	吸收(S)
脂肪酶	荧光(S)
脂肪酶(土壤)	晕圈(A)
pNB 酯酶	吸收(S)
荧光假单胞菌酯酶	pH 指示剂(A)
多黏类芽孢杆菌葡糖苷酶	颜色(A)
β-内酰胺酶(土壤)	颜色(A)
Scytalone 脱氢酶	pH 指示剂(A)
纤维素酶	晕圈,克隆大小(A)
磺基转移酶	吸收(S)
硝酸和亚硝酸还原酶	吸收(S)
限制性内切酶	FCS(S)
脱氢酶(土壤)	颜色(A)
P450 加氧酶	颜色(A)
黄色短杆菌类胡萝卜素基因簇编码的酶	颜色(A)

注:A = 琼脂板检验;S = 溶液检验;FRET = 荧光共振能量转移;FP = 荧光偏振;FCS = 荧光关联能谱法。

总之,高通量筛选是发现新酶的重要工具。定向分子进化的发展可以预期在不久的将来,将使生物催化剂及催化过程以崭新的面貌出现在生命科学和生物技术“世界”。

3.2 核糖体和 mRNA 展示技术

Mattheakes 等[5] 首先开发出核糖体展示技术[图 3.2(a)], Hanes 和 Pluck-

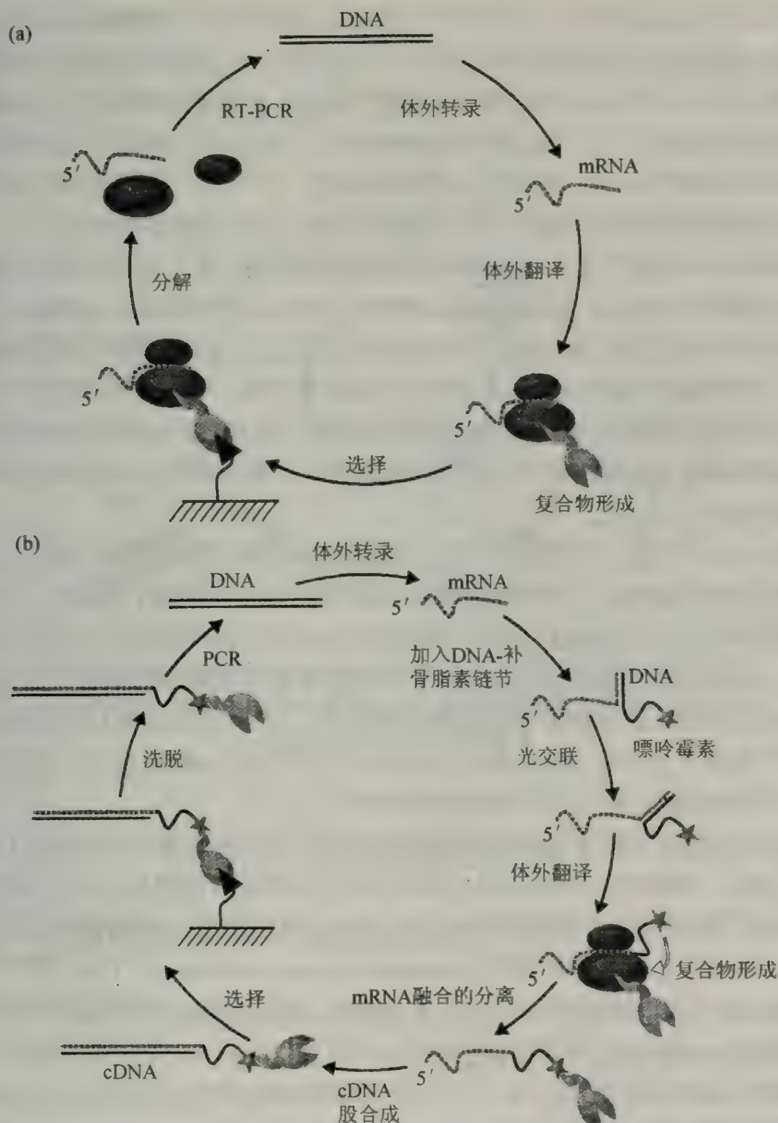


图 3.2 体外展示技术

(a)核糖体展示选择图解。编码目标蛋白质的 DNA 库与链节融合,其允许蛋白质折叠而链节仍在核糖体"隧道"中。形成的融合物(其缺乏终止密码子)体外转录为 mRNA 并进一步体外翻译,形成稳定的 mRNA,核糖体和初生多肽的三元复合物,翻译终止。这个复合物直接在固定的靶上进行结合选择。结合复合物中的 mRNA 用 EDTA 与核糖体分离,继之用 PCR 进行逆转录反应产生选择克隆。然后分析这些克隆或用作下一轮选择。(b)mRNA 展示选择图解。编码目标蛋白质的 DNA 库体外转录,形成的 mRNA 共价与短的 DNA 链节(在它的 3′端携带嘌呤霉素部位)融合。该连接可以通过杂化实现,继之补骨脂素标记的 DNA 链节与 mRNA 光交联或通过酶促连接反应,形成的连接物体外翻译。在翻译时,核糖体完成 RNA-DNA 连接和停顿,允许嘌呤部位与核糖体 A 位结合。因此,初生多肽被转移到嘌呤霉素部位导致共价 mRNA-多肽融合。这个 mRNA-融合复合物随后纯化,并且第一个 cDNA 股被合成,然后结合在固定化配基上进行选择,并且洗脱结合复合物。继之 PCR 产生选择克隆,然后分析这些克隆并用作下轮选择。图(a)和(b)中,黑线相当于 DNA,虚线相当于 mRNA。(b)中星号代表嘌呤霉素部位。为简化起见,体外翻译步骤之后交联的 mRNA-DNA 杂合体未标出

thun[6]、He 和 Taussing[7]进一步改进了折叠蛋白质的选择。核糖体展示技术是依靠 mRNA,核糖体和初生多肽的非共价三元复合物,保证基因型和表型的偶联,所构建的融合蛋白质中目标结构域与 C 端链节融合,这样该结构域可以折叠而链节仍然在核糖体“隧道”中。融合物在 mRNA 水平上缺乏终止密码子,这样阻止 mRNA 和多肽由核糖体释放,高浓度的镁和低温进一步稳定三元复合物。这个复合物是在体外翻译过程中形成的,可以直接用于展示蛋白质的各种性质。

Robert 和 Szostak[8]发展了 mRNA 展示的相关技术[图 3.2(b)],称为 mRNA 蛋白质融合或体外病毒,这个方法依靠 mRNA 与初生肽共价偶联。mRNA 首先与携带嘌呤霉素部位的短的 DNA 链节共价连接,然后这个库体外翻译,像在核糖体展示一样,当核糖体完成 RNA-DNA 连接时,核糖体停顿,并且嘌呤霉素部位进入核糖体的肽酰基转移酶位点,与初生肽形成共价链。这样,蛋白质和 mRNA 偶联,继而由核糖体分离和纯化。在目前方案中,cDNA 股被合成以形成 RNA-DNA 杂合体,其最终用于选择。

mRNA 展示方案已有明显地改进:功能 mRNA-蛋白质融合的产量与原方案比大约提高 40 倍;mRNA 与含嘌呤霉素的 DNA 链节的连接也明显改进。在新方法中,携带补骨脂素(psoralen)的 DNA 链节是同 mRNA 的末端杂化,并定向与 mRNA 光交联。mRNA 展示的这些改进打开了处理复杂蛋白质库的一个窗口。核糖体和 mRNA 展示的弱点是 mRNA 对核酸酶降解敏感。为了改进 mRNA 展示的稳定性,Kurz 等[9]报道了用双股 cDNA 取代 mRNA-蛋白质中的 mRNA 分子。该方法对于苛刻条件(例如高温)的选择可能有特殊的意义。

用定向进化技术使蛋白质亲和性成熟是产生选择和高亲和性结合蛋白质的一个重要步骤。核糖体展示和 mRNA 展示可以对各种结合的靶蛋白进行选择。

Irving[10]等报道了亲和成熟的新方法,即 Qβ 复制酶移入核糖体展示技术。这种策略可以描述为连续体外进化(continuous *in vitro* evolution, CIVE)(图 3.3)。从目标(scFv 或 V 结构域)DNA 模板出发,mRNA 在兔网织红细胞偶联系统中用 T7 聚合酶转录。mRNA 用 Qβ 酶复制和突变,同时翻译和展示:①依赖 T7DNA 聚合酶转录的模板 mRNA 编码蛋白质;②用 Qβ 复制酶复制和突变编码突变的蛋白质。在缺乏翻译终止密码子的情况下,稳定的核糖体三元复合物展示从头合成的蛋白质,并且对固定化靶(抗原)选择。编码 mRNA 是用模板特异引物的 RT-PCR 进行回收。Qβ 复制酶被加入到偶联转录翻译展示系统兔网织红细胞裂解液中,或者是纯化的 Qβ 复制酶的全酶或者是 Qβ 核心酶(β 亚基)。

对于大多数生物技术和医学应用的蛋白质的共同的要求是它们对变性剂稳定性问题。核糖体展示是体外蛋白质稳定性进化的有效工具。进化的 scFv 片段在没有二硫键的情况下是稳定的。当二硫键被改造,scFv 突变体较相应的野生型蛋白质更稳定。应用热或蛋白酶作为选择压力进行体外展示,特别是 mRNA 展示可

以提高蛋白质的稳定性。

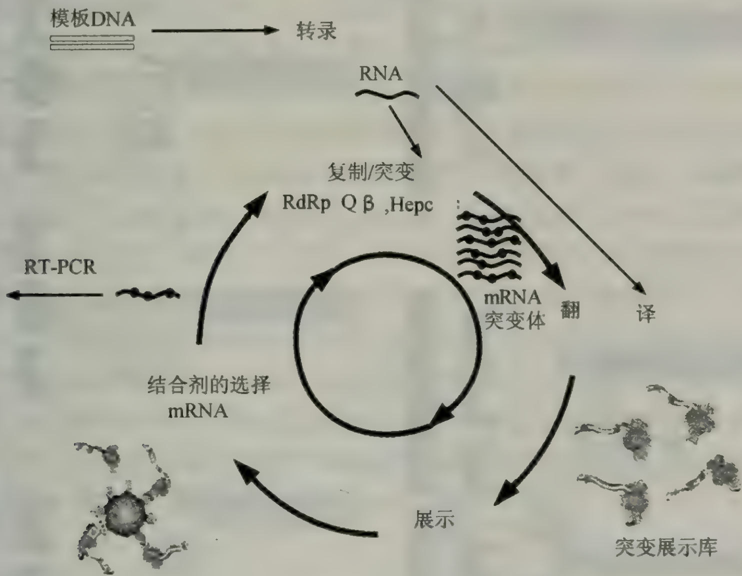


图 3.3　连续体外进化(CIVE)循环,具有 Qβ 复制酶偶联转录—翻译系统
mRNA 是通过 T7 聚合酶由 DNA 模板转录的,通过 Qβ 复制和突变,翻译并展示在核糖体的表面,然后对着靶选择

3.3　噬菌体展示技术

3.3.1　噬菌体展示作为酶库筛选的工具

酶展示主要是在丝状噬菌体表面或微生物细胞表面进行,使在目标基因和酶之间产生物理连接(图 3.4)。这个途径的明显优点是展示的酶接近内部的周围环境,这样有效地同底物或其他靶分子相互作用。

酶库筛选有不同的方式:①与过渡态类似物结合的蛋白质的选择;②根据自杀抑制剂或底物的机制富集。Olsen 等[11]报道应用生物素 N-ε-氨己酰基-胱胺-N'-戊二酰基肽加成化合物筛选枯草杆菌蛋白酶库。展示功能酶的克隆生物素化,并且可以由库中富集;③反应底物与噬菌体共价连接。酶催化噬菌体-底物或者固定化或者由固体表面释放噬菌体颗粒。这种策略依靠酶和底物邻近偶联约束噬菌体。应用邻近偶联可以分离活性酶。该途径是通过主要外壳蛋白质基因 VIII 的非特异烷基化把底物结合到噬菌体上。生物素化的 DNA 被连接到肽上,形成具有第二个 PIII 融合肽的异质二聚体超螺旋,与所展示的酶融合是不同的[图 3.4(c)][8]。

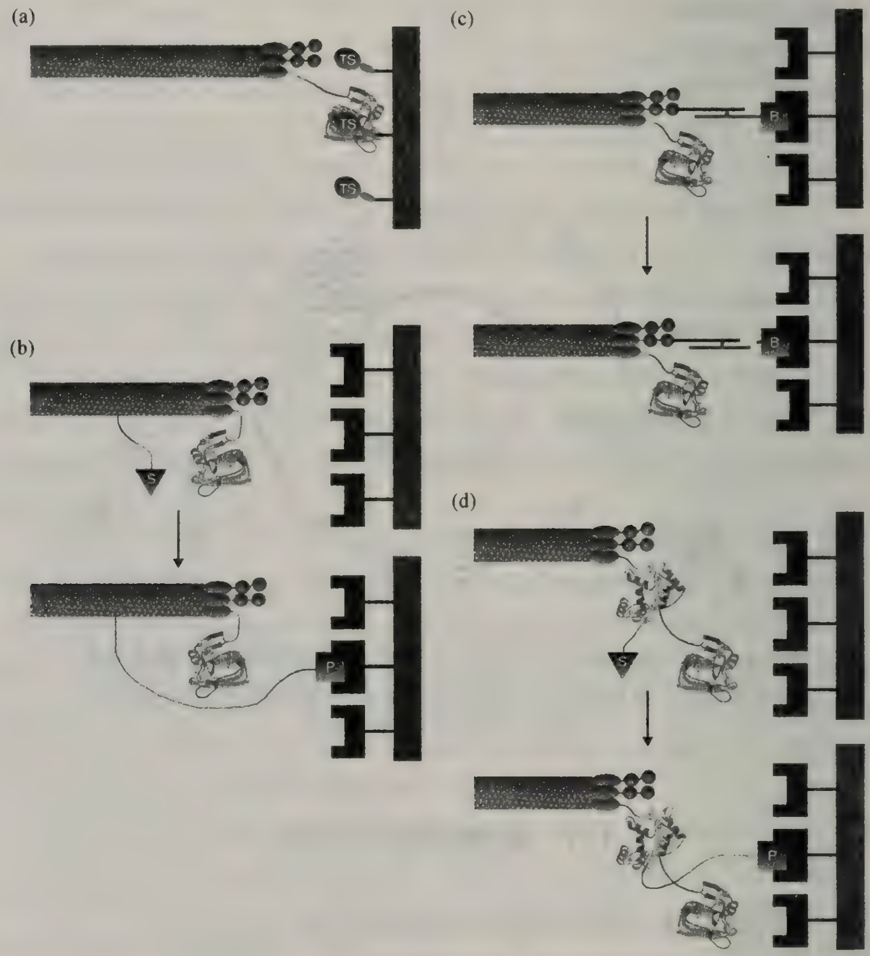


图 3.4　噬菌体展示酶库的筛选策略

(a)表达 gp^{III}-酶的噬菌体与其系在固相支持物上的过渡态类似物(TS)结合;(b)表达 gp^{III}-酶的噬菌体经 gp^{III}烷基化也展示底物(S)。在催化时,产物(P)是特异地识别并结合在固体支持物上。(c)表达 gp^{III}-酶的噬菌体及 gp^{III}-底物嵌合体,通过与底物融合的生物素标签与固相支持物结合。在催化时,噬菌体由表面释放出来。(d)底物(S)通过钙调蛋白结合的肽吸附在 gp^{III}-钙调蛋白-酶的噬菌体上。在催化时,产物(P)通过特异识别与固相支持物结合

在 PIII 融合肽和底物之间的超螺旋形成导致噬菌体固定在链霉亲和素念珠上。表达的活性葡萄球菌核酸酶水解 DNA 并选择性地由固体基质释放。

Dermatis 等[12,13]表明底物和酶展示可以组合到单一 PIII 融合形式。简言之,两侧有长的柔性链节的钙调蛋白(CAM)被插入在 PIII 蛋白和酶之间。钙调蛋白

结合的肽-底物连接物通过酶形成产物可以通过特异的抗体识别富集噬菌体。

敏感的高通量筛选进化酶，著名的例子是用脂质囊作为分离区室(一种以 3-杂合系统为基础的方法)，最终酶展示在荧光活化的细胞上(fluorescence-activated cell sorting，FACS)。Tawfik 和 Griffiths[14]探索了反相胶束水溶液核心作为体外酶合成和产物捕获的人工区室。在这个系统中，转录-翻译偶联的混合物分散在油包水乳浊液中，每个胶束平均含有一个基因。Benkovic 及其同事[11]描述了细菌三组分方法，称为 QUEST(querying for enzyme using the three-hysbrid system)。最初由 Scytalone 脱氢酶(SD)证实，如图 3.5 所示。在 SD 三组分系统中，功能转录活化剂由两个分离的蛋白质的 DNA-结合结构域二聚合产生。在这种情况下，二聚合是由合成的小分子诱导(二聚合化学诱导剂，CID)。在 QUEST 中，所设计的嵌合转录活化剂，它是 AraC 的 DNA 结合结构域同无活性 SD 突变体融合，产生 AraC′H85N-SD。两个分子的 SD 特异抑制剂连接在一起形成 CID，其可以在体内结合嵌合的转录活化剂。二聚转录活化剂调节阿拉伯糖生物合成酶的合成，导致生长介质的

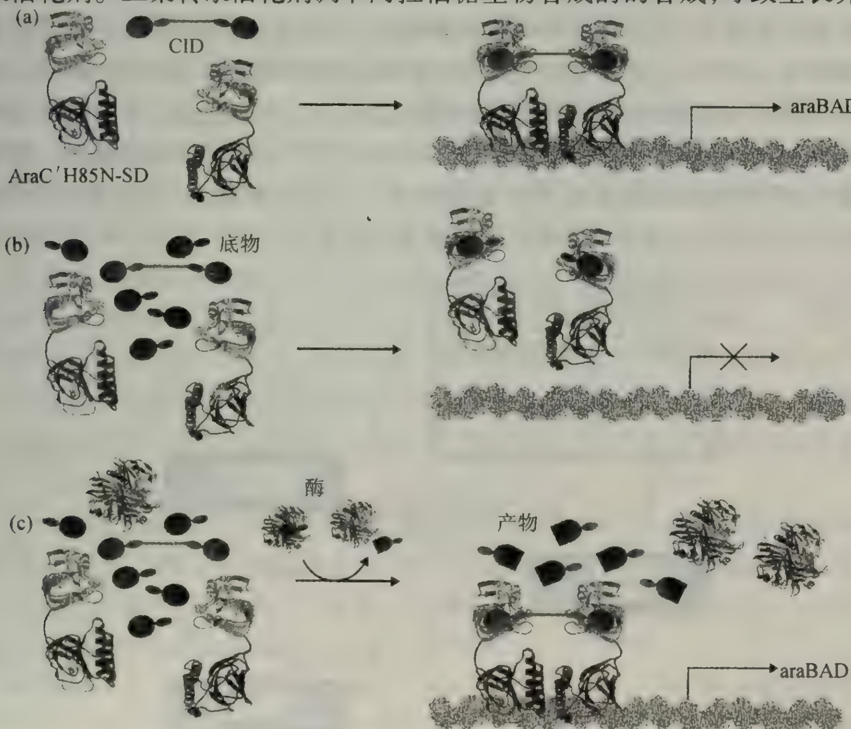


图 3.5　筛选进化酶的 QUEST 系统

(a)AraC′H85N-SD 融合蛋白与 CID 结合使 araBAD 操纵子转录；(b)底物抑制 CID 诱导的二聚合；(c)酶催化底物转为产物，CID 再次二聚合 AraC′H85N-SD，并使 araBAD 转录

酸化，其本身在平板上产生各种表型。Scytalone 与 CID 竞争，并抑制酸化表型。如果 SD 活性在细胞内被抑制，Scytalone 转化为产物，改变底物与 CID 比率。阿拉伯糖操纵子的转录一旦再次诱导，发生介质酸化。

酶展示在微生物的表面和用 FACS 筛选引起人们极大关注[16~18]。流动细胞计量术早已用于单一细胞水平分析酶活性和动力学，对大量酶的检验，包括酯酶、蛋白酶、过氧化物酶、脂肪酶和氧化还原酶等也是有效的，并且广泛用于研究和临床实践。然而，流动细胞计量术作为定向分子进化的筛选工具，在底物设计上存在的几个障碍：底物必须能够渗透到细胞中，并酶促分裂产生高荧光产物；此外，产物被细胞定量保持，底物和产物都不影响细胞生存活力；对细胞中不相关酶的修饰敏感；或表现低的溶解度等。现在，几种装备可以容易满足酶展示在微生物的表面，这样底物不再内在化，通过产物保持来拓展生物表面的物理化学性质（图 3.6）。革兰氏阴性菌的表面有大量的负电荷，聚阴离子底物用荧光共振能量转移（FRET）到 *E. coli* 的表面。底物的猝灭剂部位被表面展示的酶分裂产生荧光产物，它经聚阴离子保持在表面，荧光量与底物转换率相关。结合底物的突变酶，不提高猝灭剂的荧光强度，大概由于猝灭剂的荧光团是被酶活性部位螯合。由于这些特征，展示在表面的酶与 FRET/FACS 偶联筛选，根据 K_{cat}或 K_m 来鉴定酶。作者应用这种策略分离蛋白酶突变体，与野生型比较，在 S1 和 S1′亚位点有不同的氨基酸。流动细胞计量术，特别是与细胞表面展示偶联时有几个突出的优点：①定量分析活性；②同时可以测定并记录多个参数；③高通量（10^9/h）。

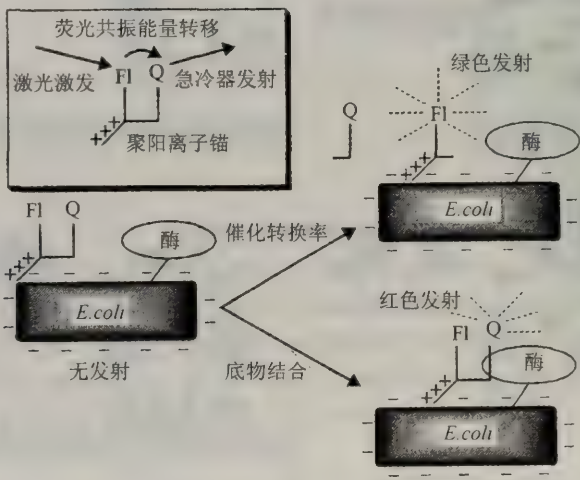


图 3.6　依赖 FACS 酶进化系统

表达高绿色荧光发射细胞是通过流动细胞计量术由库的混合物中进行物理挑选

3.3.2　噬菌体-酶

当将酶展示在噬菌体表面时，系附在噬菌体颗粒上的功能酶水平一般在突变和选择之前进行估价。这个水平等于酶和噬菌体浓度之间的比率。

纯化的噬菌体-酶溶液的噬菌体浓度可以通过测量在 265nm 处光密度来估算。吸收系数可以根据噬菌体的所有组成（氨基酸和核苷酸）的个别系数计算。g^8p 是主要贡献的蛋白质，吸收系数与噬菌体长度成正比，噬菌体与本身的单股 DNA 大小成正比。对于 10kb 噬菌体而言，它等于 8.4×10^7L/(mol·cm)。噬菌体浓度也可以通过特殊的染料和标准样测定蛋白质的含量或 DNA 的含量来估算。上述方法有可能估算过高，如果 g^3p 融合蛋白是 g^3p 主要来源的话，对于丝状噬菌体组装的正确终止来讲，将不是最佳的结果。在上述条件下，可能产生多噬菌体（含有一个以上 DNA 分子）。多噬菌体对 g^3p 蛋白质突变体已报道。感染颗粒的浓度通常是在噬菌体浓度的 1% 和 5% 之间，这是丝状噬菌体一种特征，在低水平展示时，g^3p 融合蛋白一般不影响噬菌体感染。

展示酶的浓度是通过测定噬菌体-酶溶液的活性来评价，前提是假设展示的酶像溶液中的自由酶。酶-g^3p 融合蛋白质的存在也可以通过抗-g^3p 或抗-酶抗体的 Western blotting 分析来检验。用抗-g^3p 检验，融合和自由 g^3p 蛋白质的相对强度应当与酶和噬菌体浓度之间的比率评价展示水平相一致。

很显然，噬菌体展示技术需要适当折叠的蛋白质，并且稳定地与壳蛋白融合。成功地展示可能取决于细菌宿主和生长条件。无效蛋白质折叠是一个潜在的问题，特别是展示的蛋白质不是来自细菌。真核生物蛋白质一般难在细菌中以活性形式表达。改进折叠效率的一种方式是用超表达分子伴侣工程宿主。应用该途径，Soderlind 等[19]通过 GroEL 和 GroES 分子伴侣蛋白协助在噬菌体上改进了展示抗体片段。

另一种可能降低表达水平的机制是融合蛋白质的降解，通过细菌蛋白酶降解融合蛋白质，不太稳定的突变体在噬菌体上展示的少。降低培养温度似乎是普通用于提高展示水平，降低蛋白质水解。当降低感染细菌生长温度（37℃→23℃）时，TEM-1 β-内酰胺酶展示水平提高 5 倍。

在噬菌体生产时，除蛋白质水解外，纯化的噬菌体由于蛋白酶的污染也能损失它们展示的蛋白质。应当加入蛋白酶抑制剂 Cocktail 以阻止蛋白质水解，特别是噬菌体长期贮存。

在噬菌体载体中，终止密码子往往被引入外源蛋白质和噬菌体壳蛋白质之间。因此，当感染含有适当校正基因的 *E. coli* 菌株展示才是可能的。这个系统不需要在选择后自由酶再克隆的各种步骤。

体外选择的成功不仅取决于功能展示水平，而且原则上也取决于起始库的质量和多样性。自 1991 年以来，30 种以上的各种酶已成功地在噬菌体上展示。由于在丝状噬菌体所表达的蛋白质在噬菌体组装之前必须运输到胞浆周围，因此毫不奇怪，主要展示的是天然分泌的蛋白质（蛋白酶，细菌细胞外或胞浆周围的各种酶，抗体……）。

3.3.3 噬菌体-酶的选择

3.3.3.1 结 合 选 择

噬菌体-酶的最初鉴定是检验噬菌体结合折叠酶的特异抑制剂或特异的单克隆抗体(mAb)，通过"生物选淘"(biopanning)容易做到（图 3.7）。噬菌体在固定化抑制剂（或 mAb）存在下温育，用可溶性抑制剂洗脱或改变条件（例如 pH）以降低亲和性。洗脱对吸附噬菌体数之间的比率同不相关噬菌体所观察到的背景比较，推断噬菌体展示的折叠酶。这个比率也与展示水平相关。

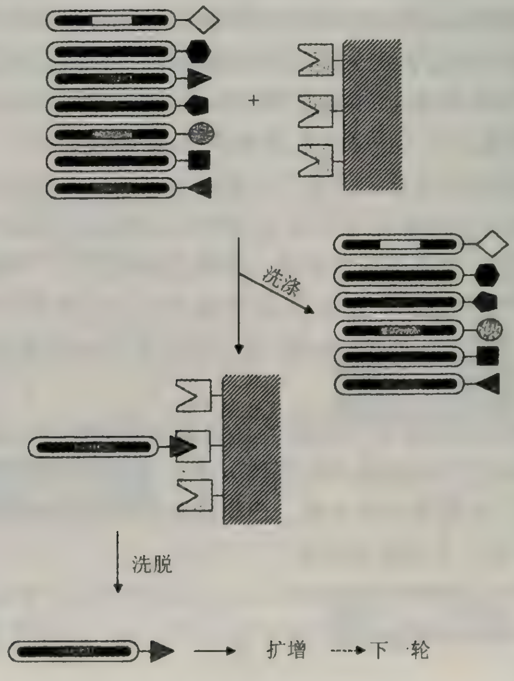


图 3.7 丝状噬菌体表面所表达的酶体外选择

在"生物选淘"实验中，特殊结合剂可以由库中捕获，并且它们的基因可以通过感染扩增

3.3.3.2　底物或产物类似物选择

底物或产物类似物抑制剂的选择已用于两种酶：葡萄球菌核酸酶和谷胱甘肽转移酶。葡萄球菌核酸酶是钙依赖的磷酸二酯酶，催化 DNA 的水解，对分裂位点 5′侧胸腺嘧啶脱氧核苷有一定的特异性。噬菌体展示葡萄球菌核酸酶突变体库是通过 epPCR 产生的，在固定化底物类似物（含胸腺嘧啶脱氧核苷或鸟嘌呤核苷磷酸硫代寡核苷酸）进行生物选淘（图 3.8），富集结合剂库。对胸腺嘧啶脱氧核苷底物类似物 **1** 所选择的最好突变体活性近于野生型酶；鸟苷底物类似物 **2** 所选择的突变体活性低于野生型 10 倍，这可能改变了特异性。

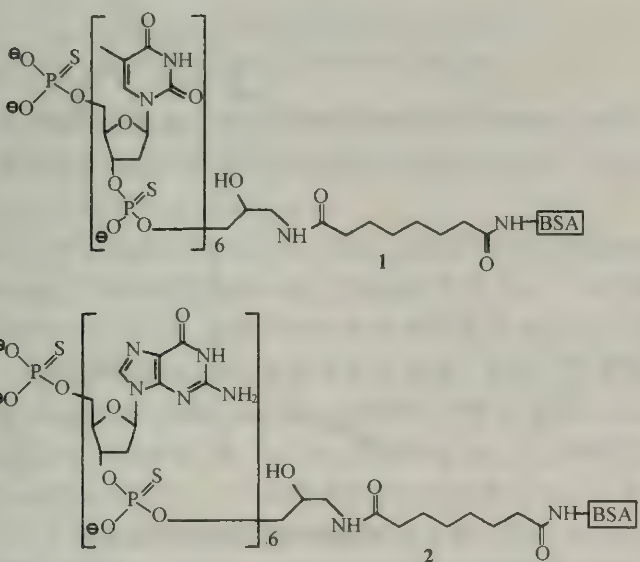


图 3.8　用底物类似物选择葡萄球菌核酸酶突变体

BSA——牛血清白蛋白

谷胱甘肽转移酶（GST）是涉及解毒过程的酶。它们催化有毒化合物和谷胱甘肽（GSH）之间的结合作为生物消除的基本步骤［图 3.9（a），**3**］。几种已知的 GST 是同源的，但它们的比活性却不同。GSTA1-1 催化谷胱甘肽 3 和氯-2，4-二硝基苯 **4** 偶联，这是亲核取代反应，改变对芳香环上携带负电荷底物的特异性。在芳香环亲电结合部位中 10 个氨基酸残基被随机突变产生噬菌体 GST 库。在类产物亲和配基上进行生物选淘，由库中选择出底物特异性改变的新 GST［图 3.9（b）-7］。

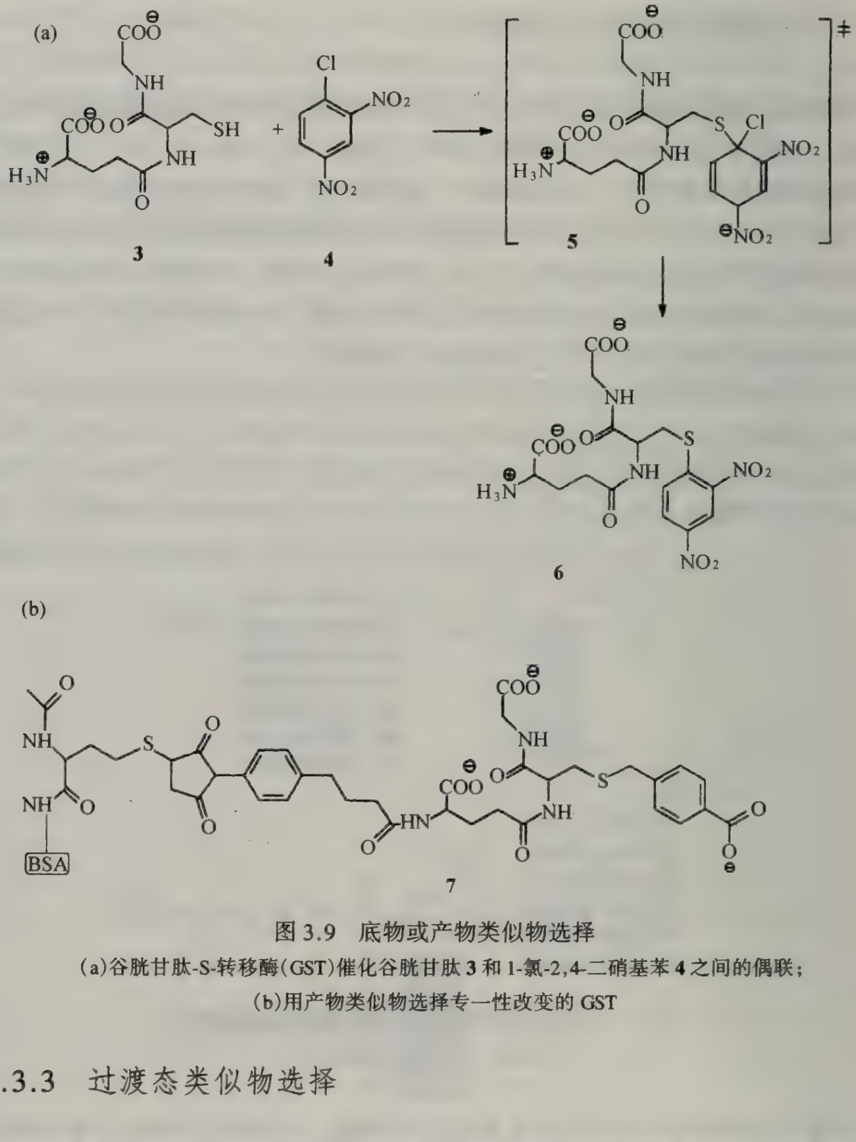


图 3.9 底物或产物类似物选择

(a)谷胱甘肽-S-转移酶(GST)催化谷胱甘肽 **3** 和 1-氯-2,4-二硝基苯 **4** 之间的偶联；

(b)用产物类似物选择专一性改变的 GST

3.3.3.3 过渡态类似物选择

根据酶催化理论，酶必须与过渡态而不是与底物基态形式互补。根据这个理论，K_{cat}/K_{uncat}等于酶对底物和过渡态的亲和常数的比率，即 K_S/K_{TSA}(图 3.10)。

$$E+S \underset{}{\overset{K^{\neq}}{\rightleftharpoons}} E+S^{\neq} \xrightarrow{\text{底物过渡态}} E+P$$

$$K_S \Updownarrow \qquad\qquad \Updownarrow K_{S^{\neq}}$$

$$E.S \overset{K_E^{\neq}}{\rightleftharpoons} E.S^{\neq} \xrightarrow{K.T/h} E+P$$

$$K_s=\frac{[E][S]}{[E.S]} \qquad K_s^{\neq}=\frac{[E][S^{\neq}]}{[E.S^{\neq}]}$$

$$K^{\neq}=\frac{[S^{\neq}]}{[S]} \qquad K_E^{\neq}=\frac{[E,S^{\neq}]}{[E.S]}$$

$$\frac{K_{cat}}{K_{uncat}}=\frac{K_E^{\neq}}{K^{\neq}}=\frac{K_S}{K_{S^{\neq}}}\approx\frac{K_S}{K_{TSA}}$$

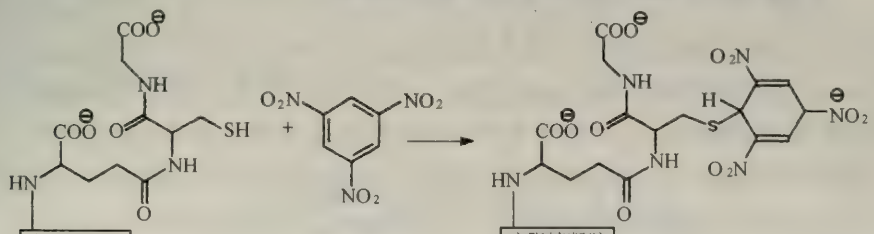
图 3.10　证实酶催化理论的热力学循环

加速速率是与酶对过渡态/底物的亲和力相关

酶这种基本识别特征导致过渡态类似物(酶的潜在抑制剂)的设计,并导致它们作为半抗原诱导免疫系统产生赋予催化活性的抗体(abzyme)。对于一个完美的 TSA 来讲,K_S/K_{TSA}可能等于 K_{cat}/K_{uncat}。

碳酸酐酶是一种金属酶,催化 CO_2 的水和,它也催化活化的羧酸酯(例如对硝基苯乙酸酯)的水解。该酶被氨磺酰抑制,后者被认为是 TSA。为了解蛋白质中金属离子亲和性的结构决定因素,直接与组氨酸金属配基相互作用的 3 个氨基酸残基用噬菌体展示的碳酸酐酶 II 的 1728 突变体库取代。用固定化氨磺酰通过它的氮与锌配位结合,层析选择锌离子亲和性,选择的突变体表明结合锌离子与野生型比较,亲和性范围从相等到降低 100 倍,80%突变体有 CO_2 水合酶活性。

过渡态类似物的亲和层析已用于由噬菌体展示库选择 GSTA1-1 突变体,亲电底物结合部位中的 4 个残基随机突变。通过 GSH 和 1,3,5-三硝基苯(TNB)之间的反应形成 σ 复合物作为亲合配体(图 3.11)。它模拟 GSH 卤代硝基苯衍生物亲核芳香取代的过渡过态,但不产生完全反应,因为它没有多余基团。在 4 轮生物选淘之后鉴定出几个突变体,其中之一由几种底物测定催化效率低于野生型 20~90 倍。



8　　9

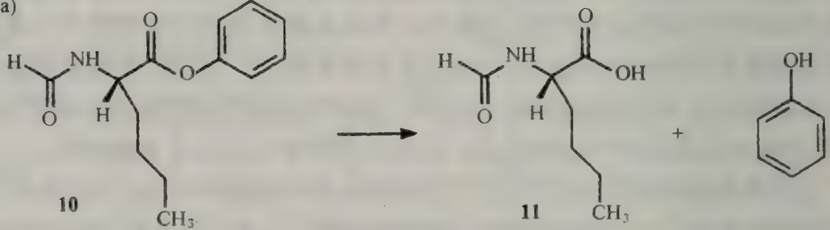
图 3.11　用固定化的过渡态类似物选择噬菌体展示的 GST 突变体[5]

突变体和野生型酶对 σ 复合物(K_{TSA})亲和力之间的比率与用 1-氯-2,4-二硝基苯所测定的 K_{cat}/K_m 值相同,与理论预测一致。

上述两个例子说明了过渡态稳定化的原理成功地用于选择活性突变体,用 TSA 选择似乎能够富集活性突变体。

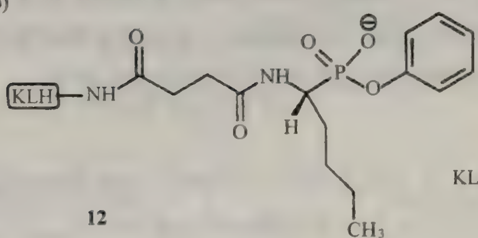
过渡态类似物选择已经用于由噬菌体展示的抗体库中选择催化抗体。Baca 等[20]用随机突变,继之膦酸生物选淘,试图提高第一代催化抗体的催化活性。从 mAb 起始,催化不同氨基酯的水解[图 3.12(a),**10**~**11**],采用"人源化"抗体以改进在 *E. coli* 中表达并适于噬菌体展示突变体的组合库进行生物选淘[图 3.12(b)]。挑选克隆进行顺序分析和鉴定。有的突变体对膦酸过渡态类似物的亲和力增加了 2~8 倍。Fujii 等[21]成功地应用 TSA 的亲和突变改进催化。用过渡过态类似物免疫产生 mAb,催化酰化的碳水化合物区域选择脱保护[图 3.13(a),**13**~**14**],速率明显提高。构建的模型指出在重链 CDR3 中 6 个氨基酸残基突变。在噬菌体展示的抗体库中这些残基随机地发生突变。从牛血清白蛋白结合过渡态类似物 15 生物选淘[图 3.13(b),**15**]和 124 随机挑选的克隆鉴定出 6 个突变体,Fab 片段以可溶形式产生。高活性的突变体 K_{cat}高于野生型 mAb 12 倍,而 K_m 增加 9 倍。有趣的是,从起始 mAb 到突变体,K_{cat}/K_{uncat}增加 12.2 倍,K_m/K_{TSA}改变与理论值基本一致。TSA 结合的最佳化可能导致 Ser-His 催化二元组匹配质量的某些损失。

(a)



10 **11**

(b)



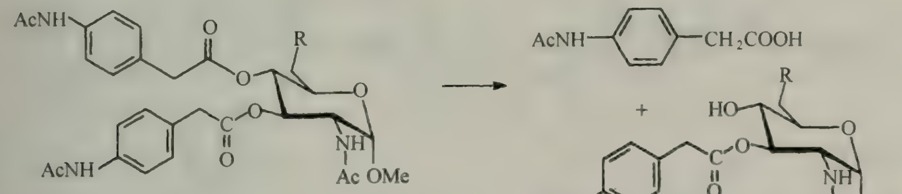
KLH = "钥匙孔"中白蓝蛋白

12

图 3.12 过渡态类似物选择催化抗体

(a) 酯酶抗体催化的反应;(b) 用过渡态类似物的衍生物进行选择

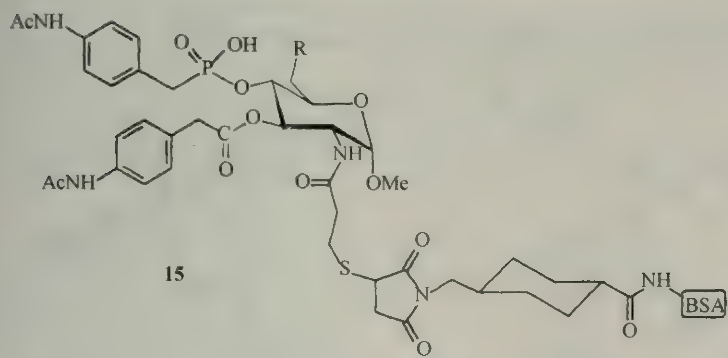
(a)



13　14

R=F 或 p-$AcNH\text{-}C_6H_4\text{-}CH_2\text{-}COO$

(b)



15

图 3.13　用 TSA 的亲合突变改进催化抗体

(a) 抗体催化区域选择酯水解；(b) 用衍生的过渡态类似物进行选择

3.3.3.4　亲和标记活性部位残基选择

酶的活性部位往往含有侧链起重要作用的氨基酸残基，通过对接活性部位设计的亲和标记试剂，能够同活性侧链形成稳定的共价键，以此由库中选择活性酶或抗体酶。

Larner 等[21,22]设计出一种试剂，能够捕获在它们的半抗原结合部位之内含有活性半胱氨酸的噬菌体-抗体。BSA-结合 α-苯乙基吡啶基二硫化物[图 3.14(a)，**16**]。该反应形成共价酰基中间体，速率提高超过背景 30 倍。虽然在酰化步骤上有限，但催化酰化半胱氨酸中间体的水解似乎更有效(相对硫代酯自发水解提高 10^4 倍)。

三羟基丁醛缩合反应是通过胺催化的，并且许多醛缩酶的活性部位含有一个主要的赖氨酸残基，应用 1,3-二酮活性免疫策略(图 3.15，**18**)，Wagner 等[23]制备了有醛缩酶活性的抗体，在其活性部位有高活性的赖氨酸残基。

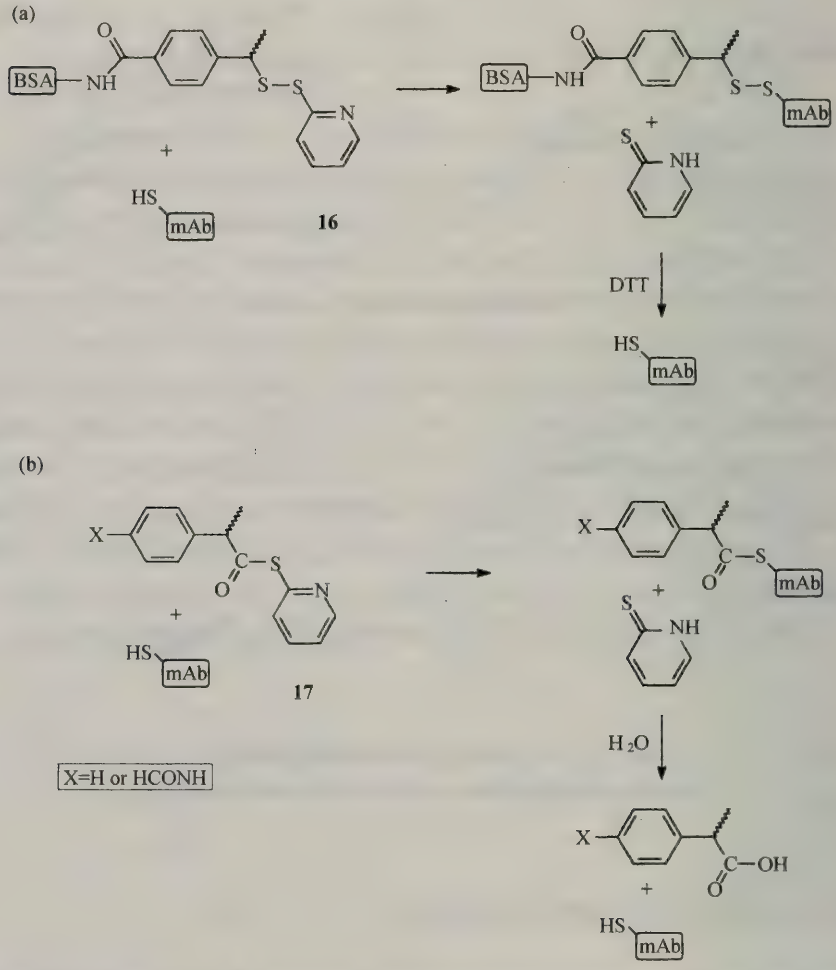


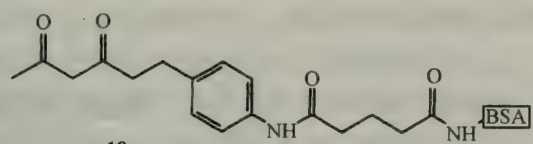
图 3.14 亲合标记活性部位残基选择催化抗体

(a) 用活性部位亲合标记试剂选择抗体结合部位的活性半胱氨酸；

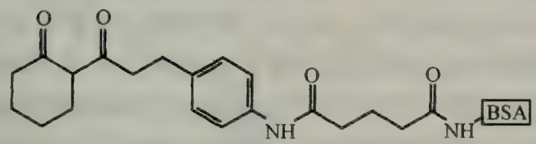
(b) 选择的抗体之一催化硫代酯水解

3.3.3.5 自杀底物选择

自杀底物使酶不可逆抑制，这是由于酶用正常机制将底物转化为抑制剂。据此，Soumillon 等[24]设计了一种选择策略，从含有无活性突变体混合物中选择活性噬菌体-酶。噬菌体展示的 TEM-1 β 内酰胺酶和无活性突变体在动力学控制下同生物素化的青霉素磺酸温育(图 3.16)。活性噬菌体-酶被优先标记，并由包盖链霉抗



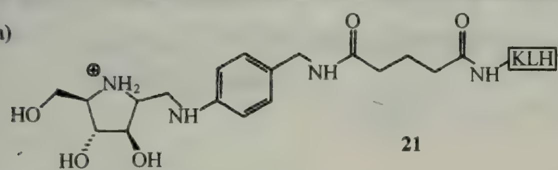
18



19

图 3.15　1，3-二酮衍生物同抗体结合部位的主要赖氨酸残基反应选择醛缩酶活性的抗体

(a)



21

(b)

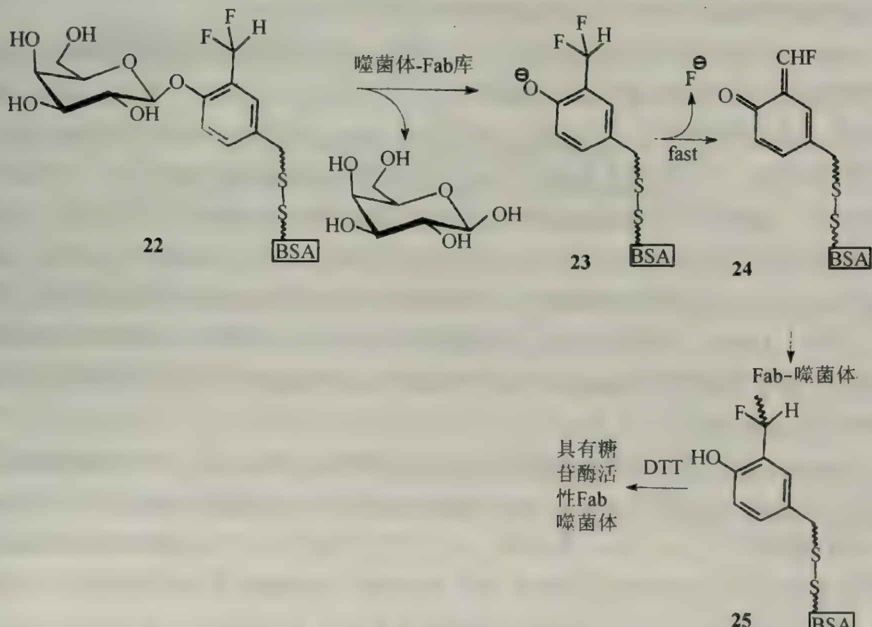


图 3.16　自杀底物选择催化抗体

(a)用过渡态类似物通过免疫产生糖苷酶活性抗体；(b)用自杀抑制剂衍生物选择噬菌体展示的抗体(Fab)

生物素蛋白的念珠结合的混合物中选择。在生物素化标记之内或在展示酶和 g^3p 之间的化学键分裂，允许回收噬菌体。该方法适于由突变体混合物中选择高活性的酶。用这种策略已由噬菌体展示库中选择出两个催化氨苄青霉素水解的催化单链抗体，K_{cat}/K_{uncat}分别为 5200 和 320(K_{cat} = 0.29/min 和 0.018/min)。

用过渡态类似物免疫诱导产生糖苷酶活性的抗体[图 3.16(a)][21]。用 BSA-缀合的自杀抑制剂(**22**)选择[图 3.16(b)]:糖苷键的分裂释放邻位二氟甲基-酚(**23**),氟化物快速消除产生非常活泼的醌甲基化物(**24**),其同噬菌体的蛋白质反应。只有催化糖苷键水解的克隆对固定化 BSA-缀合物是敏感的。噬菌体是用二硫代苏糖醇(DTT)的二硫桥使抑制剂与(**25**)中的 BSA 连接还原而释放。在 4 轮选淘之后所选择的基因再克隆,以便在 β-半乳糖苷酶缺陷株中表达可溶性 Fab 片段。涂上显色底物筛选活性 Fab 克隆,一种蛋白质已纯化,其催化对硝基苯-β-吡喃半乳糖苷水解,K_{cat} = 0.007/min 和 K_m = 0.53mmol/L,相应速率提高 $K_{cat}/K_{uncat} = 7\times10^4$。

3.3.3.6 底物转化选择

过渡态类似物往往不能足以模拟真正过渡态的本质特征。对于许多酶活性而言,没有简单的过渡态类似物或自杀底物。为了避免这些限制,提出了直接底物转换与选择过程偶联的方式,即直接根据底物转化进行选择。

Pedersen 等[26]提出的策略如图 3.17 所示,编码葡萄球菌核糖核酸酶(SNase,活性需要 Ca^{2+})的基因被克隆到噬菌粒上。为了在 g^3p 壳蛋白上展示,构建了辅助噬菌体,其加到噬菌粒上,产生展示酸性肽和 SNase 的颗粒。制备的底物的缀合物通过柔性链节,一方面与含有两个自由半胱氨酸的碱性肽顺序相连,另一方面与生物素相连。噬菌粒同底物缀合物一起温育,允许同酸性肽相互作用吸附;在噬菌粒和缀合物的半胱氨酸之间形成二硫桥以稳定缔合。这个构建物已表明在 SNase 用EDTA螯合 Ca^{2+} 而失活的条件下,与链霉抗生物素蛋白包盖的念珠有效地结合。加入可溶性 DNase I 释放全部结合的噬菌体,加入 Ca^{2+} 释放 20%结合的噬菌体。所构建的对照噬菌粒中 SNase 用 Fab 片段取代,由 SNase 和 Fab 展示噬菌粒的混合物中,可以选择前者。

Dermatis 等[12]根据底物转化提出另一选择策略(图 3.18)。用枯草杆菌蛋白酶突变体(Submut)-钙调蛋白(CaM)-噬菌体展示枯草杆菌蛋白酶突变体。Submut-CaM-噬菌体同以下 3 个肽一起温育,这 3 个肽分别含有:①N 端顺序是 Submut 水解蛋白质的产物 DYKDE;②底物 N 端有 GAAHY-DYKDE(-表示分裂位点);③负对照。所有这些肽的 C 端都含有钙调蛋白结合型主。应用识别 N 端 DYKDE 顺序的 mAb 获救噬菌体。此策略获得的酶不是完全都有活性。

用 GST-CaM-噬菌体进行的相似实验见图 3.19。底物是结合钙调蛋白的肽,

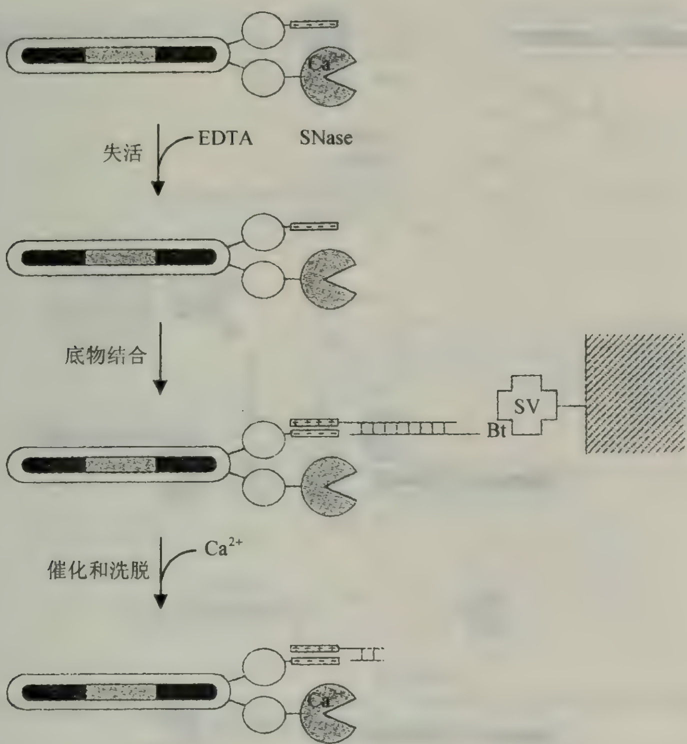


图 3.17　噬菌体展示 SNase 的选择策略
DNA 底物-端生物素，另一端有碱性肽。底物经酸性肽系
在噬菌体上，用链霉抗生物素蛋白(SV)捕获噬菌体

用其 N 端的谷胱甘肽和 4-氯-3,5-二硝基苯甲酸-生物素衍生物帽化。反应之后，噬菌体由复合物中释放，得到改进性质的突变酶。

上述选择策略的可能限制是来自过程中所涉及的分子内在性质，不足以把噬菌体结合的底物转为产物，选择的酶活性不高。为了解决这个问题，Jestin 等[27,28]设计了一种选择方案(图 3.20)。该反应方案包括酶促反应和化学反应，通过酶促反应的底物或产物标记噬菌体。用产物特异结合剂选择标记的噬菌体，只允许回收产物标记的噬菌体。

DNA 聚合酶定向分子进化引起人们极大的关注，这是因为它在生物学过程和生物技术中起核心作用。合理途径成功是有限的，这似乎由于检验的聚合酶的顺序空间有限。定向分子进化拓展了顺序空间，所希望的酶更易获得。

Xia 等[29]开发出一种选择方案(图 3.21)。他们把 DNA 聚合酶及其底物都系

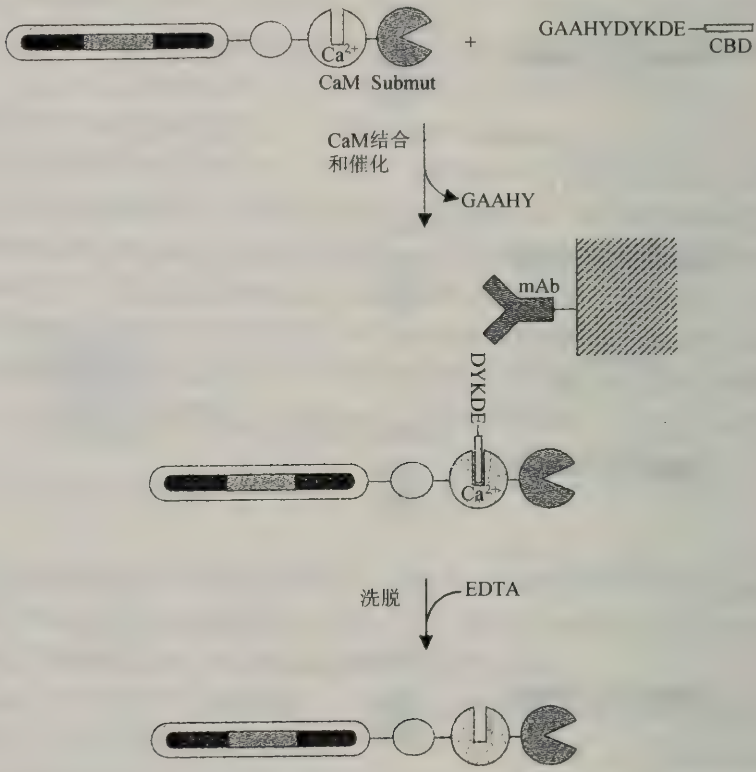


图 3.18 噬菌体展示 Submut 的选择策略

同时展示 CaM，并允许通过结合 CaM(CBD)肽捕获反应的一种产物

在噬菌体 p^{III}上。p^{III}只定位在噬菌体颗粒的一端，以便酶同底物分子反应，这样保证基因型和表型偶联。突变体的选择取决于酶催化标记的底物(例如生物素化)，这样可以选择活性聚合酶的噬菌体颗粒。Xia 等应用这个方案选择具有 RNAP 活性的 SF DNA 聚合酶突变体。分离出的 3 个突变体有效地催化底物 rNTP 的移入，其催化效率接近用 dNTP 作底物的野生型酶。

根据底物转化另一选择方案是应用催化洗脱(图 3.22)。该策略适于需要辅因子的酶，特别是金属酶。在辅因子去除后，它们仍能够结合底物。展示酶的噬菌体首先通过金属提取失活，然后它们被吸附在包盖链霉抗生物素蛋白念珠上，其生物素化的底物被固定，辅因子的加入使活性恢复并允许噬菌体洗脱，因为对产物的亲和力低于对底物的亲和力，选择效率似乎随酶在噬菌体上展示水平而提高。

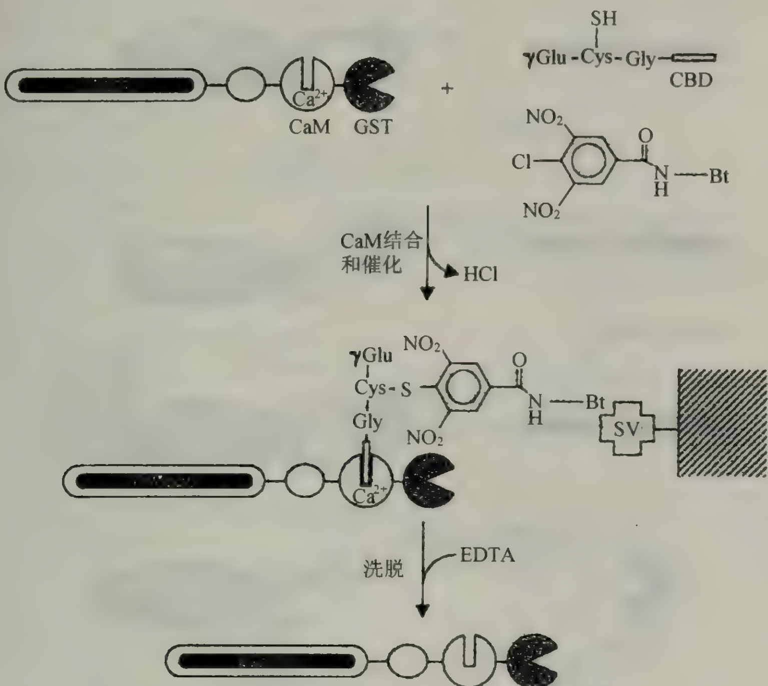


图 3.19　GST 选择策略

两种底物分别用 CBD 和生物素 Bt 衍生。含有两个模块的产物与噬菌体展示 CaM 结合，然后噬菌体用固定化 SV 捕获

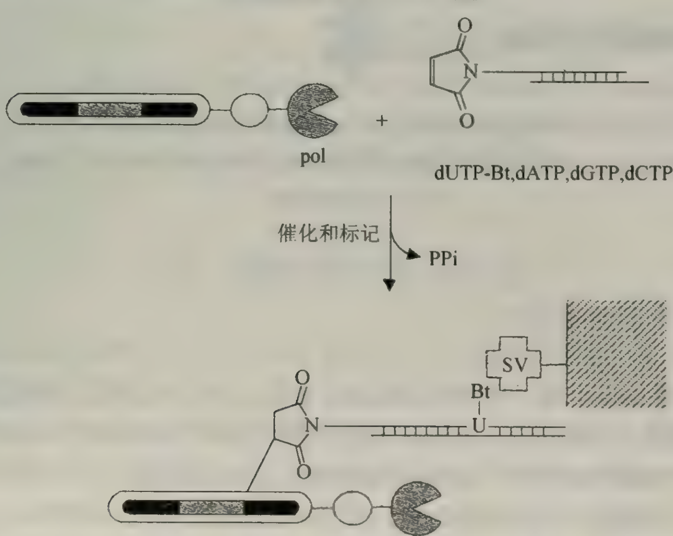


图 3.20　噬菌体展示聚合酶(pol)的选择策略

引物模板底物是用马来酰亚胺衍生，核苷酸底物用生物素(B)衍生。噬菌体经马来酰亚胺用产物标记，并用固定化 SV 捕获

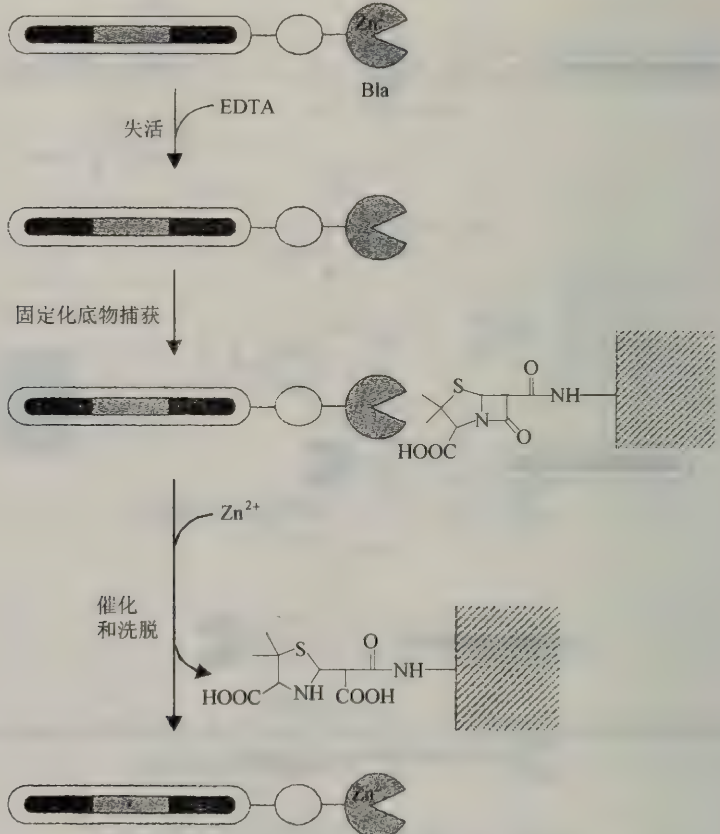


图 3.21 噬菌体展示金属-β-内酰胺酶(Bla)选择策略[24]

噬菌体-酶通过提取金属辅因子而失活，用固定化的青霉素底物捕获加入金属辅因子，产生噬菌体-酶

3.3.3.7 蛋白质折叠检验[30]

蛋白质折叠检验已用于可溶性重组蛋白质表达的定向进化。目标蛋白质的 N 端与 GFP 融合作为折叠报告。*E. coli* 细胞的荧光强度直接与可溶性蛋白质折叠的产生相关。噬菌体 F1 突变体基因 V 蛋白质，以及铁蛋白的 H 亚基在几轮定向进化之后可以用该检验方法分离出来。根据β-半乳糖苷酶的 α-片段和 ω-片段的结构互补的各种策略已被采用。目标蛋白质在 N 端与β-半乳糖苷酶的 α-片段融合，只有可溶性蛋白质可以同 ω-片段相互作用形成功能β-半乳糖苷酶，应用生色底物可以体内外检验。

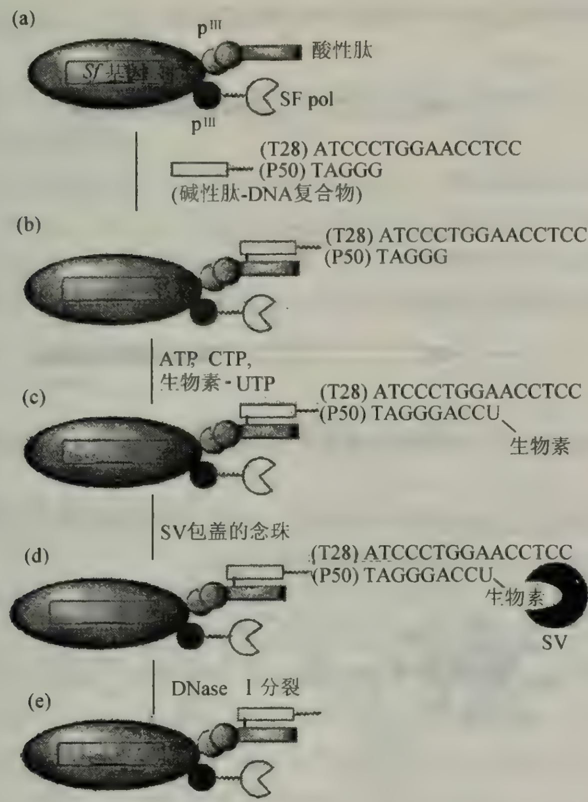


图 3.22　噬菌体选择图解

(a)展示聚合酶和酸性多肽的噬菌体颗粒;(b)核苷酸底物吸附;(c)展示活性聚合酶的噬菌体颗粒生物素化;(d)SV 珠的生物素-标签的噬菌体颗粒的固定化;(e)通过 DNase I 的裂解,固定化噬菌体释放

3.4　酵母表面展示系统[31]

目前已报道的酵母表面展示系统有两种,目的蛋白分别与 α 或 a 凝集素融合,展示于酵母细胞表面。

目的蛋白-α 凝集素表面展示系统是将目的蛋白作为 N 端,与 α 凝集素 C 端部分融合,目的蛋白经 α 凝集素展示于酵母细胞表面。α 凝集素共价连接到细胞壁的葡聚糖上,其锚定能力由蛋白质 C 端 320 个氨基酸决定,富含 Ser/Thr 残基。Ser/Thr 富集区因广泛存在的 *O*-糖基化而拥有一个杆状构象,可作为空间支撑物。

迄今,已经有多个应用 α 凝集素的 C 端作为融合蛋白的报道。第一个通过此

系统表达的异源蛋白质是 α-半乳糖苷酶(图 3.23)。

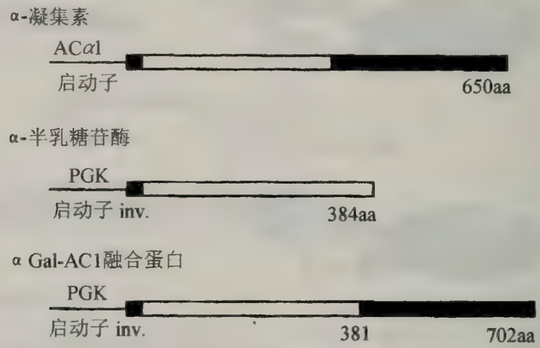


图 3.23　半乳糖苷酶-α 凝集素(α-α Gal-AC1)融合蛋白的构建

a 凝集素-目的蛋白表面展示系统是一种将目的蛋白作为 C 端与 a 凝集素 Aga2p 亚基的 N 端融合的表面展示系统(图 3.24)。

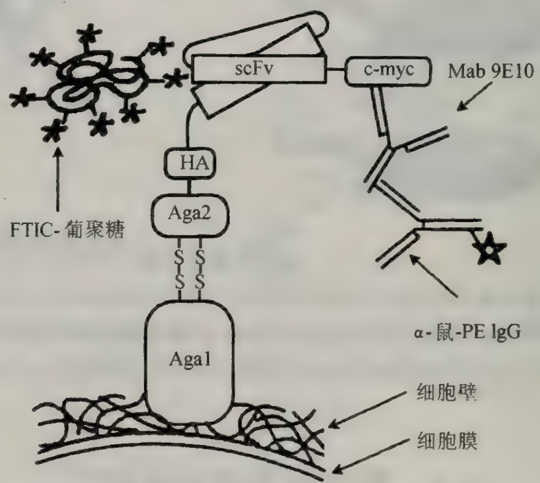


图 3.24　a 凝集素-目的蛋白表面展示系统图解

a 凝集素通过与 α 凝集素相似的连接锚定在细胞壁上。与 α 凝集素不同,a 凝集素由两个亚基的糖蛋白组成。725 个氨基酸残基的Agalp亚基通过 β 葡聚糖的共价连接而锚定在细胞壁上,69 个氨基酸残基的 Aga2p 亚基通过两对二硫键与 Agalp 亚基相连。天然的 α 凝集素结合活性部位在 Aga2p 的 C 端,因此,此处代表了一个细胞外大分子的可及性区域和展示固定蛋白的有效位点。

酵母表面展示系统的优点:①表达偏差小,DNA 库的多样性可被正确折叠蛋

白质的多样性所代表;②有定量筛选的标准;③能很好地区分微弱亲和力差别的突变体。

酵母表面展示系统用于蛋白质亲和和稳定性的定向进化已有成功报道,最先被用于抗体的亲和力成熟。

由于酵母展示的蛋白质是紧密锚固在细胞壁上,可以耐受 SDS 等的抽提。同时酵母生长快,因此在工业上具有很好的应用前景。例如,将不同特异的金属结合蛋白表达在酵母表面产生环境净化微生物,可用于处理废水中金属离子和放射性物质。将具有催化活性的酶固定在酵母细胞壁上,可防止酶的不可逆抑制,再生酶的活性。

3.5　酵母 n-杂交系统[32]

3.5.1　酵母双杂交检验

1989 年,Fields 和 Song[33]提出了"酵母双杂交检验",提供了体内检验蛋白质-蛋白质相互作用的直接方法。该法有 3 个优点:①简单、快速;②两个相互作用的蛋白质顺序可以由编码它们的质粒 DNA 顺序直接读出;③它不依赖相互作用蛋白质的特性,所以普遍适用。酵母双杂交(Y2H)检验是根据真核生物转录激活物可以分为功能上独立的结构域:DNA 结合结构域(DBD)和转录活化结构域(AD),并且杂交的转录激活物可以通过两个结构域混合和匹配产生。它似乎只需要 DNA 结合结构域将活化结构域引入转录起始位点附近。这表明 DBD 和 AD 之间的键在没有破坏活性的情况下就可以产生。这样,Y2H 检验中的键是两种相互作用蛋白质之间的非共价键。

如图 3.25 所示,Y2H 系统是由两种蛋白质嵌合体和来自转录激活物的结合位点的报告基因下游组成。如果两个目标蛋白质(X 和 Y)相互作用,它们有效地二聚合形成结合 DNA 的蛋白质嵌合体(DBD-X)和转录活化蛋白质嵌合体(AD-Y)。DBD 和转录 AD 的二聚合帮助补充转录装置,活化报告基因的转录。

最初应用两种已知体内物理联合的蛋白质:酵母 SNF1 蛋白质(丝氨酸-苏氨酸蛋白质激酶,与 GalDBD 融合)和 SNF1 激活物蛋白质 SNF4(与 Gal4AD 融合)。Gal4 结合顺序位于 β-半乳糖苷酶报告基因上游,并整合在酵母染色体的 URA3 位点上。编码蛋白质的质粒转化到酵母中,β-半乳糖苷酶合成水平应用标准生物化学定量检验。当 DBD-SNF1 和 SNF4-AD 融合蛋白质一起引入时,β-半乳糖苷酶合成水平增加 200 倍。通过比较,直接的 DBD-AD 融合蛋白质活化 β-半乳糖苷酶合成水平达 4000 倍。

自 Fieds 和 Song 文章发表以来,"酵母双杂交检验"已有明显的改进,降低了假

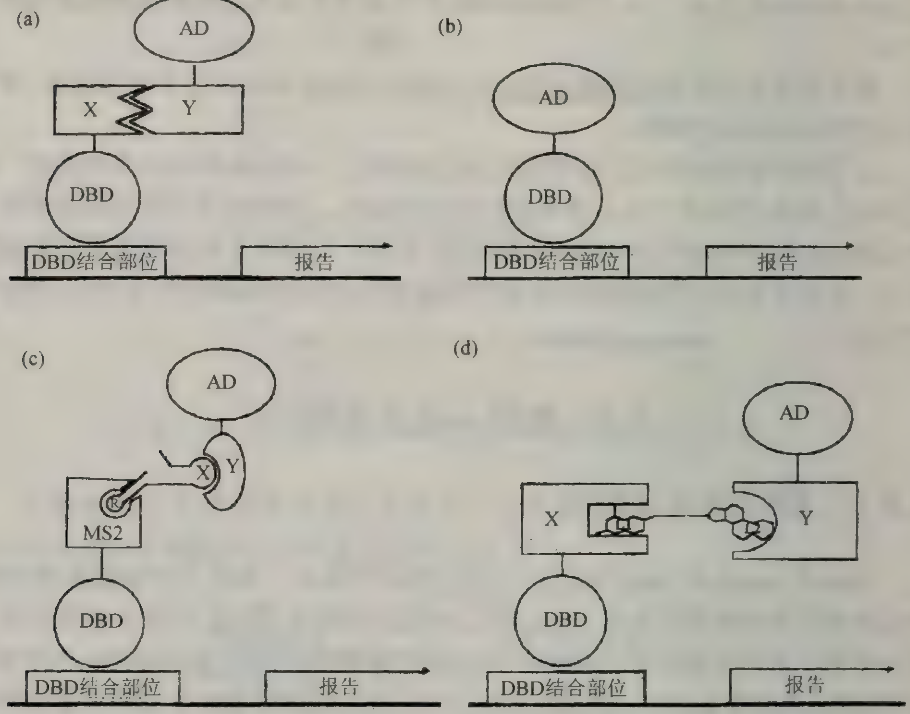


图 3.25 蛋白质-蛋白质,蛋白质-DNA,蛋白质-RNA 和蛋白质-小分子相互作用的各种酵母 n-杂交系统

(a)在 Y2H 系统中,报告基因的转录活化是通过蛋白质 X 和 Y 直接相互作用活化结构域(AD)补充重新组成。因为蛋白质 X 与 DNA 结合结构域(DBD)融合,蛋白质 Y 与 AD 融合。(b)在 Y1H 检验中,AD 直接与 DBD 融合。该检验可以用于筛选结合特异 DNA 顺序的 DBDs,或者对既定 DBD 体内结合位点。(c)同 Y2H 比较,用于检验 RNA-蛋白质相互作用的 Y3H 系统有一个杂合 RNA 分子的附加成分。一半杂合 RNA 是已知的 RNA,其可以结合 MS2 壳蛋白(MS2),具有高亲和力。另一半 RNA X,同蛋白质 Y 相互作用。(d)Y3H 可以用于检验小分子相互作用[96]。L1 是同蛋白质 X 相互作用的配基,L1 可与配基 L2 连接。这样,如 L2 同 Y 相互作用,报告基因的转录活化将重新组成

阳性,增加了转录读出。各种工程的报告质粒已广泛用于蛋白质-蛋白质相互作用的检验。现在 Y2H 检验已用于鉴定上千种新的蛋白质-蛋白质相互作用。同样的特征使其成为筛选 cDNA 库的积极方法,并用于定向进化(图 3.26),也在尝试检验 DNA-蛋白质相互作用。酵母单杂交(Y1H)和检验 RNA-蛋白质,小分子-蛋白质相互作用的酵母三杂交可能有一系列应用前景。n-杂交检验已成功应用于蛋白质进化。

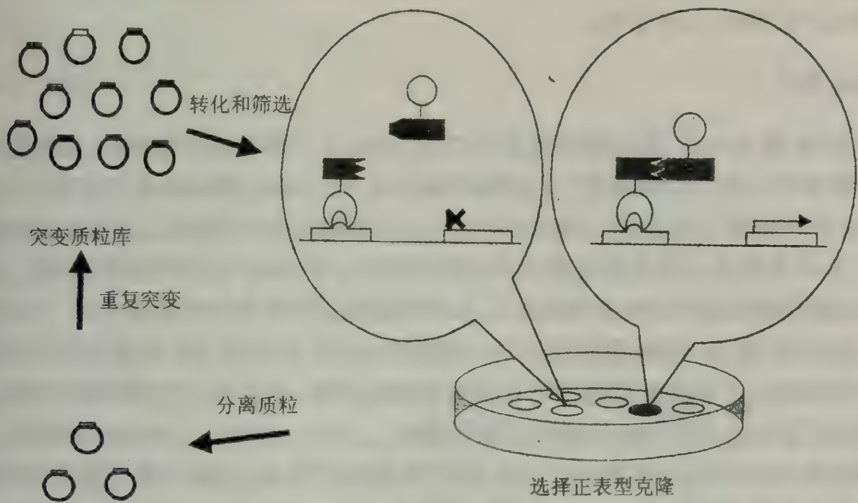


图 3.26　Y2H 检验为进化蛋白质提供了简便的方法

如图所示，第一步携带与 AD 融合的蛋白质基因的质粒在特异区域突变产生蛋白质突变体库。质粒被转达化到宿主，它是含有编码蛋白质-DBD 融合以及报告质粒的菌株，其有依赖于两种蛋白质相互作用确定的模型，含有相互作用的细胞变为蓝色用于选择。编码 AD 融合蛋白质的质粒从这些蓝色细胞中分离。然后这些质粒或者顺序分析，或进行下几轮突变和选择，以提高蛋白质-蛋白质相互作用的亲和力

3.5.2　酵母双杂交系统

自 1989 年提出 Y2H 系统以来，基本系统已做了某些改进。各种结合 DNA 和活化结构域已做了介绍。具有各种细菌耐药标准的常规载体，酵母的复制起点和酵母营养缺陷型标记已商业化。普通酵母遗传标记已引入 Y2H 系统，应用生长选择，使其能够检验蛋白质突变体大库（ca10^6）。除了基本激活物系统，逆转和分离-杂交系统，已用于检验蛋白质-蛋白质相互作用的破坏。基于转录阻遏物系统已报道。本节也探讨了 Y2H 选择出现的假阳性的本质及如何减少假阳性，真核生物中转录机制，蛋白质-蛋白质相互作用如何决定 Y2H 系统中转录活化的水平等问题。

3.5.2.1　结合 DNA 和转录活化结构域

自从第一个 Y2H 系统报道以来，几个实验室集注在方法学的最佳化研究，并已改进了一些基本系统。最大的改进是使该系统更适于直接由 cDNA 库检验蛋白质-蛋白质相互作用，这些变化也证明其在分子进化应用的有效性。下面介绍的几

个商业上有效的 Y2H 系统。

1. Gal4 系统

Fields 和 Song[33]提出的系统是由 Gal4 DNA 结合结构域(DBD)和 Gal4 活化结构域构成的。两个结构域置于乙醇脱氢酶启动子 ADH1* 截短变体的控制之下,其是一个中等强度的启动子。编码这些结构域的质粒是穿梭载体,这样可以在酵母和 *E. coli* 中繁殖。两个质粒都含有 pBR 细菌 Ori 和 Amp^R,以及高拷贝酵母 2μ Ori 和营养缺陷型标记 TRP1 和 LEU2,以选择性的保持酵母中的质粒。

过去 10 年,对 Gal4 系统进行了一些修饰,以使其适合于广泛范围的蛋白质。因为某些蛋白质在高浓度表达时对酵母细胞有毒性,于是用具有着丝粒(CEN)Ori 的质粒代替 2μ。CEN Ori 保持质粒拷贝数在 1 ~ 5 个/细胞(2μ Ori, ca60/细胞)。另一种改进是应用全长 ADH1 启动子,较之截短的突变体,它指导蛋白质更高表达。为了促进各种蛋白质和 cDNA 库的次克隆,在 Gal4 *DBD* 和 *AD* 基因的 N 端和 C 端引入多克隆位点,允许在 3 个读框中以正反方向实现 N 端和 C 端融合。

任何 Y1H 筛选的最后一步是分离编码目标的质粒,或者顺序分析或者进一步突变和选择。在典型的 Y2H 实验中,酵母含有编码 DBD 和 AD 蛋白质融合的质粒,常常编码报告基因第 3 或第 4 个质粒或第 3 个蛋白质。这样,就会产生如何选择编码进化蛋白质的质粒的技术问题。一种途径是并入可选择的反向标记,例如将 URA3 并入所有质粒,除了编码选择的蛋白质突变体的质粒外,URA3 还编码乳清酸核苷-5′-磷酸脱羧酶。该酶不仅是尿嘧啶生物合成途径中的一种主要酶,而且也是非毒性 5-氟乳清酸(5-FOA)转为毒性 5-氟尿嘧啶(5-FU)的一种酶。这样,含有 URA3 标记可以对着 5-FOA 选择。或者,该质粒可以由酵母分离出,并直接转到 *E. coli*。然后,在 *E. coli* KC8($trp1^-$ 和 $leu2^-$)分离出所要求的质粒,或者编码进化的蛋白质突变体的质粒工程到含有惟一抗生素标记,例如 Amp^R、Kan^R、Cam^R 或抗 *Zeocin* 基因(Zeo^R)菌株中。Mumberg 等[34]为 Gal4 系统设计的质粒,其允许 Ori(2μ 或 CEN)、启动子(ADH1, CYC1, TEF2 或 GPD)以及营养缺陷型标记(HIS3, LEU2, TRP1 或 URA3)进行交换。

2. GAl 系统

改进的 Gal4 系统是采用 GAl 启动子。它调节酵母中的启动子,并且可以用半乳糖诱导高水平表达,完全被葡萄糖抑制,以及随着半乳糖和葡萄糖比率变化而调节。当然这个启动子不能用于 Gal4 系统,因为 Gal4 是 GAl 启动子系统的一部分。控制蛋白质融合的表达特别适于分子进化。当早期几轮选择时,蛋白质可以高水平表达。为了辨别高活性的蛋白质突变体,在后几轮选择时蛋白质的表达水平可能降低。在诱导剂缺乏的情况下根据活化报告基因筛选诱导系统,可迅速检验出

假阳性。

3. LexA 系统

在 Gal4 DBD 位置，采用了 *E. coli* 的 $LexA_{202}$ DBD，并且 GAl 上游活化顺序(UAS)用多拷贝 LexA 操纵子取代。在 Gal4 AD 位置，该系统的 Vojtek 变体是采用疱疹单显性组合病毒 VP16 蛋白质的 AD。LexA 系统与最初 Gal4 Y2H 非常相似。这样，许多成分可以交换，对 DBD 和 AD 质粒来讲，Vojtek 系统应用同样的营养缺陷型标记，分别是 TRP1 和 LEU2，如 Gal4 系统。它应用 ADH1 启动子控制 $LexA_{202}$ 融合蛋白质的表达，并且 ADH1 启动子控制 VP16 融合物的表达，这些质粒也有 2μ Ori，主要差别是应用非常强的 VP16 活化结构域。VP16 AD 同样含有 SV40 的 T 抗原核定位顺序(NIS)，Gal4 AD、$LexA_{202}$ DBD 不含有 NIS。这种设计使核中 AD 融合蛋白质的浓度大于 $LexA_{202}$ 融合蛋白质，增加 $LexA_{202}$ AD 复合物结合报告基因的上游的机会。LexA/VP16 系统的突变体已经发展。

3.5.2.2　报告基因

Y2H 系统特别适于分子进化实验，因为它可以进行生长选择。此外，应用生色报告基因(例如 *LacZ*)的次级筛选可以排除假阳性。在早期几轮筛选中，可以应用敏感的报告基因检验活性的蛋白质。在后几轮，应用更多的选择性报告基因辨别高活性的突变体。Y2H 报告质粒提供 3 个控制水平——上游活化顺序(UAS)、启动子和报告基因本身。报告质粒的选择是最佳敏感度和可接受背景之间的一种平衡。

UAS 提供第一个控制水平。UAS 通过 DBD 所识别的 DNA 区域，并且 DBD-AD 复合物与 UAS 的结合负责转录活化。已确定转录活化水平与 DBD 结合位点数相关，这样可以制备具有 1～4 串联结合位点的一系列报告基因。

启动子提供第二个控制水平。应该强调，除了活化转录外，还存在启动子顺序支配基础转录的问题。报告的敏感度可以通过转录的基础水平来提高。它也是最佳敏感度和可接受背景之间的一种平衡。

报告质粒的第 3 个成分是报告基因本身。已发展了一些报告系统。Y2H 检验可以用比色测定筛选正或负生长。最初 Y2H 系统应用 *LacZ* 基因以提供蛋白质-蛋白质相互作用的比色测定。*LacZ* 基因编码 *E. coli* β-半乳糖苷酶，其有广泛的底物特异性，可以分裂一些生色底物，包括 5-溴-4-氯-3-吲哚-β-D-吡喃半乳糖苷(X-gal)，邻硝基苯-β-D-吡喃半乳糖苷(ONPG)。它们可以在平板上或液体培养中迅速检出。除了 *LacZ*，几个其他的比色测定报告基因也是有效的。*MEL1* 和 *gusA* 分别编码 α-半乳糖苷酶和 β-糖苷酸酶，可以用生色底物 5-溴-4-氯-3-吲哚-α-半乳糖苷

(X-α-gal)和 5-溴-4-氯-3-吲哚-β-D-葡萄糖醛酸(X-gluc)来检验[35]。编码绿色荧光蛋白(GFP)的基因已在 Y2H 系统作为比色测定报告,并且可以不用加入任何外源底物进行检验。

3.5.2.3 基础双杂交系统的修饰

Y2H 系统的修饰使其可以检验蛋白质-蛋白质相互作用的破坏。应用反选择性标记检验蛋白质-蛋白质相互作用的破坏[图 3.27(a)]。反向 Y2H 系统应用 *Gal4 DBD*、*AD* 及 *URA3* 报告基因。首先,蛋白质-蛋白质相互作用是根据在缺乏尿嘧啶介质中生长来确定。其次,破坏相互作用的分子可以根据 URA3 转录的破坏

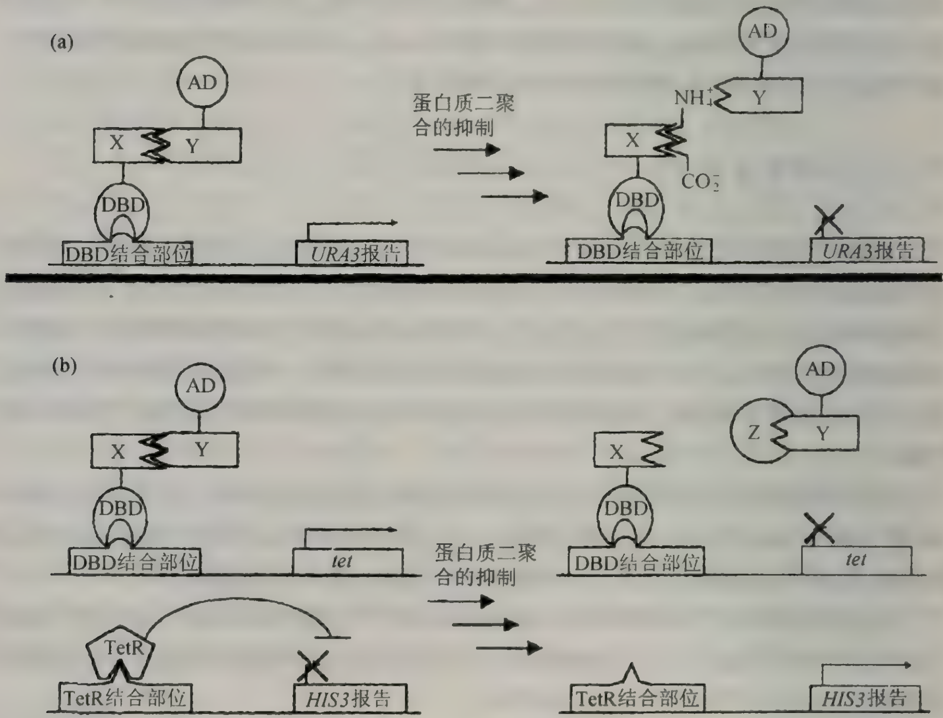


图 3.27 检验蛋白质-蛋白质相互作用破坏的 Y2H

(a)反向 Y2H 系统,X 和 Y 之间蛋白质-蛋白质直接相互作用,活化 *URA3* 转录,其将 5-FOA 转为毒性化合物(左);然而,如果 X 和 Y 之间相互作用被破坏,*URA3* 转录被破坏,细胞在含 5-FOA 介质中生长(右)。(b)在断裂-杂交系统中,X 和 Y 之间的蛋白质-蛋白质相互作用活化 *tet* 转录,*tet* 能编码细菌的结合 DNA 的蛋白质 TetR。TetR 依次结合 *HIS3* 报告基因的上游,并阻止该基因的转录,其对于缺乏组氨酸介质中生长是主要的(左)。在 X 和 Y 之间这种相互作用的破坏,通过第 3 个蛋白质 Z 关闭 *tet* 基因的转录。在缺乏 TetR 的情况下,*HIS3* 基因不再被抑制,允许细胞在缺乏组氨酸介质中生长

和在 5-FOA 存在下生长来检验。

在断裂-杂交系统中，蛋白质-蛋白质相互作用的破坏是应用双报告系统来检验[图 3.27(b)][36]。蛋白质-蛋白质相互作用指导 *E. coli* Tet 阻遏蛋白基因 *tet* 的转录(TetR，一种结合 DNA 的蛋白质)。*HIS3* 基因本身是在两个串联 TetR 操纵子控制下。该系统是根据带有 $LexA_{202}$ DBD 和 VP16AD 的 Vojtek 系统。最初蛋白质-蛋白质相互作用活化 TetR 转录，然后 TetR 通过基础转录机制在物理上阻断转录来抑制 HIS3 的产生。这样，在组氨酸缺乏情况下根据生长来选择，它已用于破坏蛋白质-蛋白质相互作用的蛋白质的进化。

另一系统是研究蛋白质-蛋白质相互作用抑制转录激活物系统(图 3.28)，所谓抑制转激活物系统(RTS)是根据 Gal4 Y2H 系统，只是 Gal4 AD 是用酵母阻遏蛋白 TUP1 的 N 端抑制结构域(RD)取代[37]。该系统是为初始负生长选择设计的，采用 *URA3* 报告继之 *LacZ* 筛选。*URA3* 基因是在 *GAl1* 启动子控制之下。这样，如果没有蛋白质-蛋白质相互作用发生，*URA3* 基因被活化，并且酵母对 5-FOA 敏感。然而，如果 Gal4 DBD 和 TUP1 RD 融合蛋白质相互作用，那么 TUP1RD 被充实到 *GAl1* 启动子，并且抑制 *URA3* 转录，允许在 5-FOA 存在下生存。

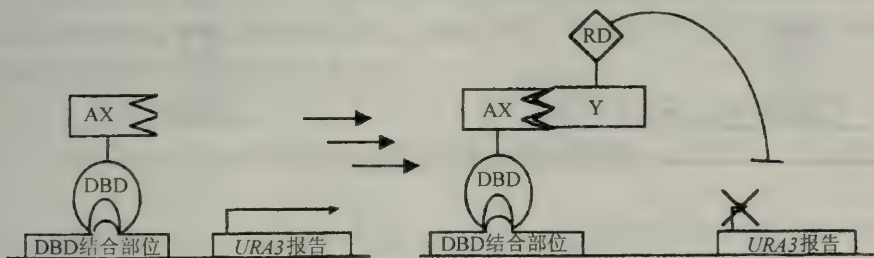


图 3.28 抑制转录激活物系统

该系统可以用于独立活化转录的各种蛋白质。因为被筛选的蛋白质 AX(活化的蛋白质 X)有内在活化 *URA3* 报告基因转录的能力，AX 的表达使细胞对 5-FOA 敏感(左)。TUP1 的抑制结构域(RD)的补充，通过指导活化蛋白质 AX 和蛋白质 Y 之间的相互作用，抑制 *URA3* 基因的转录，允许细胞在含有 5-FOA 介质中生长(右)

关于降低在 Y2H 检验中观察到的假阳性问题已进行了详细地讨论[37]。降低假阳性的关键是在开始选择之前挑选最佳的 Y2H 系统，排除假阳性最普通的方式是应用不同的报告质粒进行实验，或二级筛选。

3.6 细菌双杂交(B2H)系统

在细菌中工作有几个潜在的优点,对分子进化实验来讲,基本的优点是*E. coli*的转化效率高于酵母几个数量级,一次可检验大约 10^8 蛋白质突变体(酵母是 10^6)。另外,分子生物学技术已在 *E. coli* 中最佳化,选择时间缩短。细菌双杂交系统示如图 3.29。

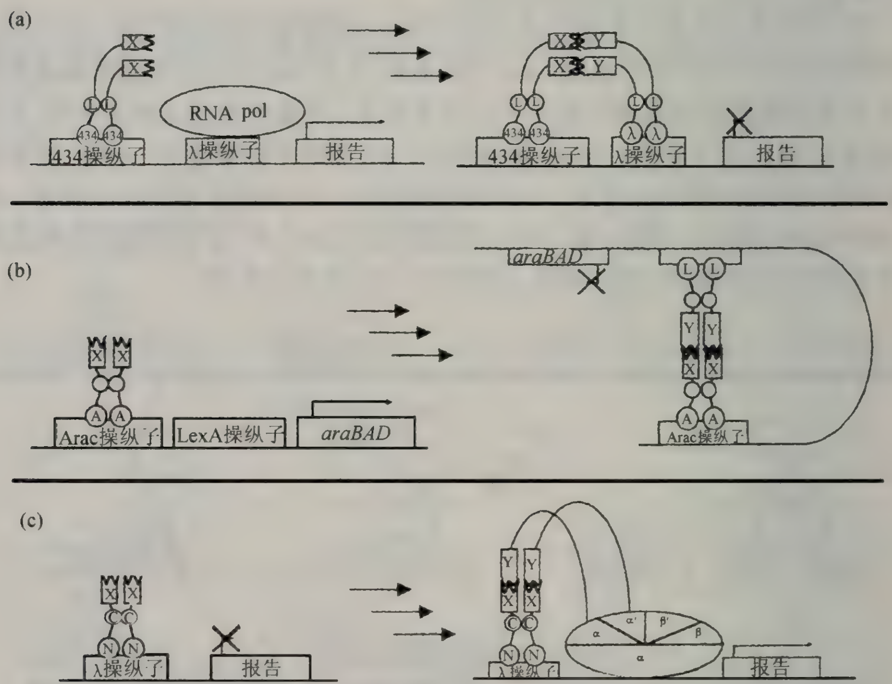


图 3.29 细菌双杂交(B2H)系统

(a)最初的 B2H 系统。它是根据细菌的天然阻遏蛋白。图中蛋白质 X 是与 434 阻遏蛋白亮氨酸(L)拉链的 C 端融合,其与 434 操纵基因结合。434 操纵基因的存在不影响 RNA pol 的结合或报告基因的转录(左)。如果蛋白质 X 同蛋白质 Y 相互作用,那么 λ-阻遏蛋白融合补充低亲和力 λ 操纵基因,阻止 RNA pol 与启动子区域结合,从而抑制报告基因的转录(右)。(b)AraC-LexA 系统。它是根据抑制 DNA 结合。AraC-蛋白质 X 嵌合体活化 Ara BAD 操纵子(左),串联的 *LexA* 操纵基因工程在 Ara BAD 操纵子的转录和翻译起点之间。AraC-X 和 LexA-Y 蛋白质的相互作用引起真实的 DNA 结合和转录抑制(右)。(c)细菌激活物系统。λ-阻遏蛋白通过稳定 RNA pol-启动子复合物简单地活化转录。这样,如果蛋白质 X 和 Y 相互作用,蛋白质 X-λ 阻遏蛋白融合和蛋白质 Y 与 RNA pol 的 α-亚基的 N 端结构域融合,活化报告基因的转录

许多细菌阻遏物和激活物是二聚体蛋白质，在结构上具有不同结合 DNA 和二聚化的结构域[图 3.29(a)]。λ-阻遏蛋白 C 端二聚化结构域可以用酵母转录激活物 GCN4 的亮氨酸拉链二聚化结构域取代。该嵌合蛋白质是稳定的，在功能上可以取代体内 λ-阻遏蛋白。围绕 LexA 和 AraC 发展了相似的系统，并且可以检验均一和异质二聚体[38]。应用 AraC-LexA 系统改进了 *LacZ* 抑制的动态范围[图 3.29(b)]。在该系统中，3 个串联的 LexA 操纵子已工程在 AraCBAD 启动子控制的 *LacZ* 报告基因的转录和翻译的起点之间。AraC 的表达或 AraC-融合蛋白活化该启动子的转录。如果 LexA-融合蛋白质表达，其可以同 AraC-融合蛋白质相互作用，然后相互作用产生启动子 DNA 弯曲。这个弯曲阻止 *LacZ* 基因转录，大概破坏 RNA 聚合酶的结合或活性。在该系统中，*LacZ* 转录水平可提高 100 倍，这取决于蛋白质-蛋白质相互作用强度。

一个细菌激活物系统如图 3.29(c)。它是简单地把 λ 操纵子置于转录起始位点的上游。蛋白质-蛋白质相互作用可以根据 λ-阻遏蛋白-融合蛋白稳定 RNA pol 同启动子相互作用能力来检验。在该系统中，Gal4 同 Gal11 相互作用使 *LacZ* 转录提高 45 倍。

细菌双杂交系统的开发只是刚刚起步，难点是在细胞中难于表达多个质粒，并且染色体整合也不像在酵母中重要。因此，根据不同的应用，工程含有报告基因的菌株，对于细菌系统将是特别重要的。

3.7　蛋白质互补检验

检验蛋白质-蛋白质相互作用的其他方法，一般地依赖于诱导蛋白质的两个互补片段相互作用，造成酶活性或 FRET 的重建。根据两个蛋白质片段的一种主要酶的重建，几个实验室已设计出高通量检验方法。很明显，酶的选择是关键。所需要的酶可以分成两个片段，且必须不能自身缔合，即使在超表达的情况也应如此。片段的二聚合必须保持适当的折叠和功能。为了选择和筛选，酶必须提供明显的表型。一种这样的检验是根据腺苷酸环化酶合成腺苷 3′,5′-环化单核苷酸(cAMP)。*B. pertussis* 的腺苷酸环化酶可以分为两个功能上互补的片段(T18 和 T25)，根据 T18 和 T25 的二聚化检验蛋白质的二聚化，重建腺苷酸环化酶活性部位。因为 cAMP/CAP 复合物诱导几个基因的转录，包括 Lac 操纵子，应用 β-半乳糖苷酶平板检验或在乳糖上生长来筛选腺苷酸环化酶的活性，为此需要敲除该酶的菌株。同样，围绕 β-半乳糖苷酶的 α-亚基和 ω-亚基的二聚化和二氢叶酸还原酶(DHFR)的 N 端(1 ~ 105)和 C 端(106 ~ 186)片段设计出新的系统。

3.8　酵母 n-杂交系统的应用

3.8.1　蛋白质-蛋白质相互作用

Y2H 检验首先用于分子进化，鉴定具有高亲和性结合目标蛋白质的各种肽。由 Gal4 AD 融合的随机肽库选择出视网膜细胞瘤蛋白（Rb 蛋白）肽配体。应用 HIS3 生长选择，继之用 *LacZ* 筛选，结合 Rb 的 7 个肽已由 3×10^6 转化体库中分离。Y2H 检验也用于鉴定肽适体。

双杂交实验不局限于肽，也用于产生具有新特异性的 PDZ 结构域。B2H 检验也用于进化抑制蛋白质二聚合的肽。

3.8.2　蛋白质-DNA 相互作用

为了检验 DNA-蛋白质相互作用，可以直接"one-hybrid"检验。RNA pol 检验进化特异结合 DNA 的锌指突变体。在该检验中，蛋白质-蛋白质相互作用是根据结合 DNA 的蛋白质和 RNA pol 的 α-结构域的二聚合以及依赖 RNA pol 转录活化进行检验。为了读出蛋白质-DNA 相互作用，3 个串联的锌指被融合到 Gal11，其以高亲和力与 Gal11 结合，与 RNA pol 的 α-结构域融合。此外，如果锌指以高亲和力与所要求的 DNA 顺序结合，它们可以活化 *HIS*3 报告基因的转录。

3.8.3　蛋白质-RNA 相互作用

RNA-蛋白质相互作用的选择不太直接，因为 RNA-蛋白质融合不能直接在体内产生，常规生化检验困难。但这可以通过第 3 种成分加入到双杂交系统按 Y3H 检验（图 3.27）。第 3 种成分是一种杂交 RNA 分子，一半是已知的 RNA，它与已知的蛋白质以高亲和性结合；另一半是目标 RNA 分子。即 Y3H 系统是由两个蛋白质嵌合体，一个 RNA 嵌合体和报告基因组成。一种 RNA Y3H 系统可以检验两个已知的蛋白质-RNA 配对之间的相互作用。应用 RNA Y3H 系统筛选 RNA 分子随机库可以鉴定未知的 RNA-蛋白质相互作用。

3.8.4　蛋白质-小分子相互作用

小分子常常称为二聚合的化学诱导剂（CID），目前采用了单聚和二聚小分子。二聚 CID 更像杂合 RNA 分子那样方式应用，一半分子作为锚，另一半不确定。蛋

白质-小分子相互作用一个重要的开发领域是新的 CID 配对。

Y3H 检验应当作为工程新的小分子特异性蛋白质，B2H 检验用于蛋白质-小分子相互作用，只有少数细菌三杂交系统做了报道，这些系统仍然需要最佳化。为了用于分子进化，一个重要的步骤是设计更好的细菌小分子三杂交系统。

双杂交检验将为体内分子进化实验提供方便的工具。双杂交系统也是检验 cDNA 库和蛋白质工程的有力工具。因为检验是在体内进行，编码蛋白质突变体的质粒在每轮选择后容易回收，或者顺序分析，或者进一步突变和选择。细菌双杂交系统实验可在 *E. coli* 中进行，可检验 10^8 库。双杂交检验用于蛋白质进化是非常鼓舞人心的，可以在单轮选择中鉴定出高亲和力结合剂。三杂交系统扩展了检验，甚至将这些检验转为酶催化检验[39]。

3.9　环境库的构建和筛选

微生物统治生物圈。据计算 1g 土壤含有 4000 种不同的细菌。目前估计地球上有 1 000 000 ~ 100 000 000 种微生物，但记载的约 5000 种。由此可见，自然界蕴藏着巨大遗传信息；自然界是新基因、新产物和生物合成途径无限的宝库。

由环境样品分离新酶经典或繁琐的途径是根据要求的活性来富集和筛选各种微生物，然后酶及其基因由鉴定的微生物回收。传统培养技术需要将环境样品中各种微生物在适当的生长介质中培养，然后分离，直到分离出个别的克隆。用这种方法，由于在纯培养中富集和分离微生物的困难，能导致环境中大部分微生物多样性丧失。人们估计自然界可观察到的超过 99% 的微生物应用标准技术是不能培养的。为了解决这个问题，已开发出一些方法。该途径是应用一定环境中的微生物的遗传多样性作为整体以发现未知基因和基因产物。为了探索各种环境的遗传多样性的一种方法是由不能培养的微生物分离 DNA，继之用 PCR 扩增并克隆所要求的基因。所用的各种引物的顺序是来自已知基因或蛋白质族的保守区，以微生物生态学研究中所采用的 *16SrRNA* 基因顺序作为来源是有效的。但是为生物技术目的，完整新基因或基因产物用 PCR 鉴定是有限的。其他的方法是应用分离的环境 DNA 构建 DNA 库，并且克隆来自然样品的功能基因，不需要顺序信息，现存环境库可以用于各种目标筛选。该方法已成功获得 4-羟基丁酸脱氢酶[40]、脂肪酶[41]、酯酶[42]、淀粉酶[43]和几丁质酶[44]。构建环境库所必需的步骤及应用库鉴定新基因或基因产物示如图 3.30。关键步骤包括环境 DNA 的分离和纯化，获得的 DNA 连接到载体上，继之，由构建的环境库筛选所需基因。

样品的收集是 DNA 分离和分析的关键。对于陆地生态系统（例如土壤或微生物簇），取样不太困难。样品体积小并立即贮存冰上或冷冻。对于其他的地方，例如深海沉积和温泉收集，以及高纯度取样的地方，需要对收集、贮存或在取样处加

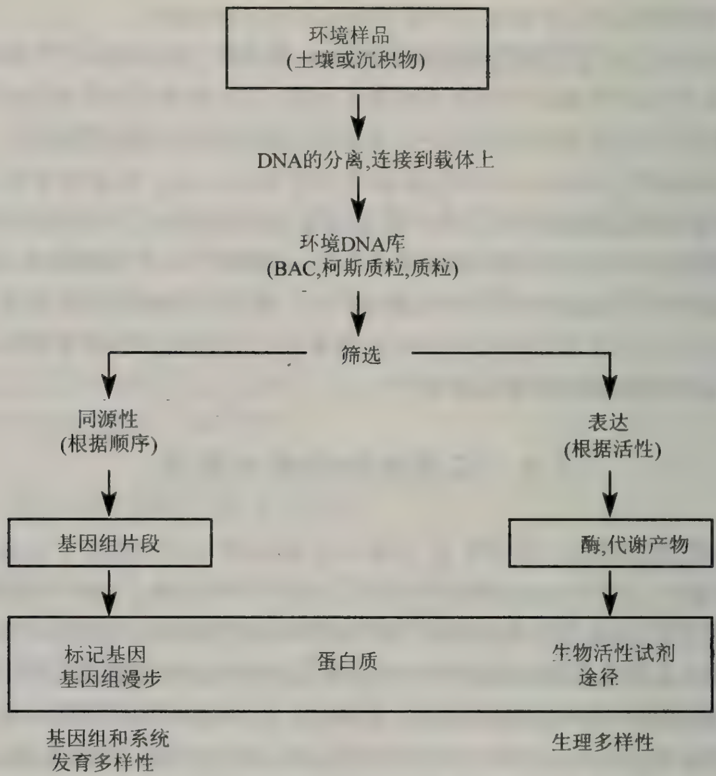


图 3.30 环境库的构建和筛选

工样品的附加工作。为了避免微生物多样性的丧失,样品运输和贮存必须小心。在贮存时,在样品冰冻和溶解之前,容易造成核酸的丢失。一般冰冻的样品在 2h 之内保持微生物的最广谱。

天然微生物功能基因的克隆取决于提取和纯化的环境 DNA 的质量。在克隆各步骤过程中所需要的酶促修饰对存在于环境生态系统中的生命和非生命成分的污染是敏感的。例如土壤 DNA 的提取总是伴有腐殖质,其干扰限制酶消化和 PCR 扩增,并降低转化效率和 DNA 杂化的特异性。为解决污染问题,已开发出一些方法。这些方法或者根据回收或者依靠微生物细胞分级和继之裂解,或样品中的细胞直接裂解,继之纯化 DNA。

直接裂解方法已得到很大发展,用高生物量环境样品,例如表面土壤或每克固体材料总细菌量 $\geqslant 10^8$ 细胞的污泥均可以分离环境中的 DNA。DNA 提取操作包括高温、去污剂处理、机械破坏、冰冻或解冻,典型的方法示如图 3.31。这种方法适于各种环境样品,包括土壤和沉积物。由环境样品提取 DNA 是在 SDS、十六烷基三甲基胺溴化物(CTAB)和蛋白酶 K 存在下用高盐提取缓冲液裂解细胞来完成

的。最终步骤是粗提 DNA 的纯化和通过阴离子交换树脂或硅胶层析去除共提取物。最终 DNA 纯化的其他方法,例如氯化铯平衡密度离心,凝胶过滤和琼脂糖凝胶电泳等。操作的主要部分是由土壤和沉积物样品回收细菌 DNA,同时无细胞 DNA 也被提取。直接裂解技术较分级方法更常用,因为它适于各种土壤和其他环境,劳动强度小,并且 DNA 收率高。由各种土壤样品回收 DNA 的量(以土壤计)2.5～26.9μg/g。从直接裂解途径分离的 DNA 可以大概代表环境样品的微生物多样性,因为不需要由环境基质回收微生物,粘到颗粒上的微生物也不易被裂解。这样,构建的库可以包含整个遗传多样性。此外,分离的 DNA≥25kb,适于所有分子操作,例如用限制酶消化和连接到载体上等。

环境样品

↓

细胞裂解:在 CTAB 和蛋白酶 K 存在下用
高盐提取制备粗 DNA,然后用 SDS 处理

↓

在 65℃培育 2～3h

↓

氯仿/异戊醇提取

↓

用乙醇沉淀 DNA

↓

通过硅胶或阴离子交换柱层析
去除共提取物

图 3.31　通过直接裂解途径由土壤和沉积物分离环境 DNA[45～48]

分级方法涉及在细胞裂解之前由环境基质分离微生物细胞。该途径对于低生物量(例如水样)或干扰 DNA 分离物质含量高(例如污水)是有优势的。一般由环境基质分离细胞是以完整样品的缓冲液匀浆开始,然后基质颗粒和其他碎片用低速离心除去,继之,未吸附的微生物细胞通过高速离心和裂解收集。为了提高 DNA 收率,可进行多轮匀浆和离心。对单一有机体而言,通过标准方法分离 DNA,最终 DNA 纯化按直接裂解途径进行,这样几乎可以由细菌获得完整的 DNA,因为细菌是来自环境基质,任何无细胞 DNA 或较大和较重的真核细胞均可以被分离出。通过分级方法回收环境 DNA 似乎不太容易被基质化合物或土壤中腐殖质污染。此外,分离的 DNA 的平均大小大于直接裂解途径所得。但是,在土壤和沉积物样品的情况下,分级途径可以优先从快速生长的细菌细胞中回收 DNA,这种选

择性会造成分离出的 DNA 遗传多样性的丧失。

在环境 DNA 库的构建中,*E. coli* 作为产生和保持环境库的宿主,因为各种遗传工具对它都是有效的;质粒、柯斯质粒、fosmid、λ 噬菌体和细菌人工染色体(BAC)作为载体。适当载体的选择取决于库所要求平均插入 DNA 的大小和拷贝数。此外,分离环境 DNA 的质量影响载体系统的选择。在某些情况下,经纯化之后,环境 DNA 仍然被腐殖质或基质污染,只有通过质粒载体克隆。质粒(例如 pBluescrip 或 pET)的另一优点是在 *E. coli* 表现出高拷贝数,允许根据活性筛选。这是非常重要的,因为环境基因的活性表达往往取决于天然启动子。该方法局限于单一基因或小操纵子的筛选,不能进行大基因簇编码的代谢活性的筛选。为了克隆环境 DNA 的大片段,应用柯斯质粒、fosmid 和 BAC。环境柯斯质粒和 fosmid 库的平均插入大小是 20～40kb,BAC 库是 27～80kb(表 3.2)。应用 BAC 载体构建的环境库已引起人们的关注。BAC 是修饰质粒,其包含来自 *E. coli* uF 因子的复制起点。BAC 的复制是被严格的控制的,保持每个细胞一或两个拷贝的复制子。这种载体可稳定的保持和复制至 600kb。由此可见,BAC 是回收大的环境 DNA 片段的优良载体,它可以发现大的基因簇和鉴定环境的系统发育多样性。BAC 库和 cosmid 库的缺点是载体的拷贝数低。从环境库中筛选所要求的基因和基因簇是鉴定新的生物催化剂、天然产物和新的分子结构的一种有效方法。环境库的筛选可以根据重组株的表型或酶活性或者核苷酸顺序进行,例如由 286 000 克隆分离出编码酯酶的 3 个新基因(表 3.2)。

表 3.2 根据活性筛选环境库的实例

DNA 来源	载体	检验的 *E. coli* 克隆数	筛选靶	正克隆数	文献
土壤	质粒	286 000	酯酶	3	[42]
土壤	质粒	930 000	4-羟基丁酸脱氢酶	5	[40]
沉积物	λ	75 000	几丁质酶	9	[44]
土壤	BAC	3 648	DNase	1	[43]
土壤	BAC	3 648	抗菌活性	1	[43]
土壤	BAC	3 648	脂肪酶	2	[41]
土壤	BAC	3 648	淀粉酶	8	[43]
土壤	BAC	24 576	溶血活性	29	[43]

注:BAC 是细菌人工染色体。

参 考 文 献

1 沈倍奋. 分子文库. 北京:科学出版社,2001

2 吴乃虎. 基因工程原理. 第二版(上册). 北京: 科学出版社,2000

3 李育阳. 基因表达技术. 北京:科学出版社,2002

4 陆德如,陈永青. 基因工程. 北京:化学工业出版社,2002

5 Mattheakes L C, Bhatt R R, Dower W J. An *in vitro* polysome display system for identifying ligands from very large peptide libraries. Proc Natl Acad Sci USA, 1994, 91: 9022 ~ 9026

6 Hanes J, Plückthun A. *In vitro* selection and evolution of functional proteins by using ribosome display. Proc Natl Acad Sci USA, 1997, 94: 4937 ~ 4942

7 He M, Taussig M U. Antibody-ribosome-mRNA(ARM)complexes as efficient selection particles for *in vitro* display and evolution of antibody combining sites. Nucleic Acids Research, 1997, 25: 5132 ~ 5134

8 Roberts R W, Szostak J W. RNA-peptide fusions for the *in vitro* selection of peptides and proteins. Proc Natl Acad Sci USA, 1997, 94: 12297 ~ 12302

9 Kurz M, Gu K, Al-Gawari A, *et al*. cDNA-protein fusions: covalent protein-gene conjugates for the *in vitro* selection of peptides and proteins. ChemBiochem, 2001, 2: 666 ~ 672

10 Irving R A, Coia G, Roberts A, *et al*. Ribosome display and affinity maturation: from antibodies to single V-domains and steps towards cancer therapeutics. J Immunol Methods, 2001, 248: 31 ~ 34

11 Takahashi T T, Austin R J, Roberts R W. mRNA display: ligand discovery, interaction analysis and beyond. Trends in Biochemical sciences, 2003, 28: 159 ~ 165

12 Demartis S, Huber A, Viti F, *et al*. A strategy for the isolation of catalytic activities from repertoires of enzymes displayed on phage. J Mol Biol, 1999, 286: 617 ~ 633

13 Fernandez-Gacio A, Uguen M, Fastrez J. Phage display as a tool for the directed evolution of enzymes. Trends in Biotechnology, 2003, 21: 408 ~ 414

14 Tawfik D S, Griffiths A D. Man-made cell-like compartments for molecular evolution. Nat Biotechnol, 1998, 16: 652 ~ 656

15 Firestine S M, Salinas F, Nixon A E, *et al*. Using an AraC-based three-hybrid system to detect biocatalysts *in vivo*. Nat Biotechnol, 2000, 18: 544 ~ 547

16 Daugherty P S, Chen G, lverson B L, *et al*. Quantitative analysis of the effect of the mutation frequency on the affinity maturation of single chain Fv antibodies. Proc Natl Acad Sci USA, 2000, 97: 2029 ~ 2034

17 Daugherty P S, Iverson B L, Georgiou G. Flow cytometric screening of cell-based libraries. J Immunol Methods, 2000, 243: 211 ~ 227

18 Holler P D, Holman P O, Shusta E V, *et al*. *In vitro* evolution of a T cell receptor with high affinity for peptide/MHC. Proc Natl Acad Sci USA, 2000, 97: 5387 ~ 5392

19 Soderlind E, Lagerkvist A C S, Duenas M, *et al*. Chaperonin assisted phage display of antibody fragments on filamentous bacteriophages. Biotechnology, 1993, 11: 503 ~ 507

20 Baca M, Scanlan T S, Stephenson R C, *et al*. Phage display of a catalytic antibody to optimize affinity for transition-state analog binding. Proc Natl Acad Sci USA, 1997, 94: 10063 ~ 10068

21 Fujii I, Fukuyama S, Iwabuchi Y, *et al*. Evolving catalytic antibodies in a phage-displayed combinatorial library. Nat Biotechnol, 1998, 16: 463 ~ 467

22 Gao C, Mao S, Lo C H, *et al*. Making artificial antibodies: a format for phage display of combinatorial het-

erodimeric arrays. Proc Natl Acad Sci USA, 1999, 96: 6025 ~ 6030

23 Wagner J, Lerner R A, Barbas C F. Efficient aldolase catalytic antibodies that use the enamine mechanism of natural enzymes. Science, 1995, 270: 1797 ~ 1800

24 Soumillion P, Jespers L, Bouchet M, *et al*. Selection of beta-lactamase on filamentous bacteriophage by catalytic activity. J Mol Biol, 1994, 237: 415 ~ 422

25 Vanwetswinkel S, Avalle B, Fastrez J. Selection of beta-lactamases and penicillin binding mutants from a library of phage displayed TEM-1 beta-lactamase randomly mutated in the active site omega-loop. J Mol Biol, 2000, 295: 527 ~ 540

26 Pedersen H, Holder S, Sutherlin D P, *et al*. A method for directed evolution and functional cloning of enzymes. Proc Natl Acad Sci USA, 1998, 95: 10523 ~ 10528

27 Jestin J L, Kristensen P, Winter G. A method for the selection of catalysis using phage display and proximity coupling. Angew Chem Int ED, 1999, 38: 1124 ~ 1127

28 Ponsard I, Galleni M, Soumillion P, *et al*. Selection of metalloenzymes by catalytic activity using phage display and catalytic elution. ChemBiochem, 2001, 2: 253 ~ 259

29 Xia G, Chen L J, Sera T, *et al*. Directed evolution of novel polymerase activities: mutation of a DNA polymerase into an efficient RNA polymerase. Proc Natl Acad Sci USA, 2002, 99: 6597 ~ 6602

30 Kawasaki M, Inagaki F. Random PCR-based screeningfor soluble domains using green fluorescent protein. Biochem Biophys Res commun, 2001, 280: 842 ~ 844

31 郭波,谢佩蓉,邹强,郑萍. 酵母表面展示系统研究进展. 生物化学与生物物理进展,2002,29:19 ~ 22

32 Brakmann S, Johnson K.(eds), Direetion Molecular Evolution of Proteins, Wiley-VCH, 2002,127 ~ 157

33 Fields S, Song O. A novel genetic system to detect protein-protein interactions. Nature, 1989, 340: 245 ~ 246

34 Mumberg D, Muller R, Funk M. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene, 1995, 156: 119 ~ 122

35 Remy I, Michnick S. Visualization of biochemical networks in living cells. Proc Natl Acad Sci USA, 2001, 98: 7678 ~ 7683

36 Michnick S W. Exploring protein interactions by interaction-induced folding of proteins from complementary peptide fragments. Curr Opin Struct Biol, 2001, 11: 472 ~ 477

37 Subramaniam R, Desveaux D, Spickler C, *et al*. Direct visualization of protein interactions in plant cells. Nat Biotechnol, 2001, 19: 769 ~ 772

38 Dimasi N, Pasquo A, Martin F, *et al*. Engineering, characterization and phage display of hepatitis C virus NS3 protease and NS4A cofactor peptide as a single-chain protein. Protein Eng, 1998, 11: 1257 ~ 1265

39 Korn K, Foerster H H, Hahn U. Phage display of RNase A and an improved method for purification of phages displaying Rnases. Biol chem, 2000, 381: 179 ~ 181

40 Firestine S, Salinas F, Nixon A, *et al*. Using an AraC-based three-hybrid system to detect biocatalysts *in vivo*. Nat Biotechnol, 2000, 18: 544 ~ 547

41 Henne A, Daniel R, Schmitg R A, *et al*. Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. Appl Environ Microbiol, 1999, 65: 3901 ~ 3907

42 Porteous L A, Armstrong J L. Recovery of bulk DNA from soil by a rapid, small-scale extraction method. Curr Microbiol, 1991, 22: 345 ~ 348

43 Henne A, Schmitg R A, Bomeke M, *et al*. Screening of environmental DNA libraries for the presence of genes

conferring lipolytic activity on *Escherichia coli*. Appl Environ Microbial, 2000, 66: 3113 ~ 3116

44 Rondon M R, August P R, Bettermann A D, *et al*. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl Environ Microbiol, 2000, 66: 2541 ~ 2547

45 Cottrell M T, Moore J A, Kirchman D L. Chitinases from uncultured marine microorganisms. Appl Environ Microbial, 1999, 65: 2553 ~ 2557

46 Zhou J, Bruns M A, Tiedje J M, *et al*. DNA recovery from soils of diverse composition. Appl Environ Microbiol, 1996, 62: 316 ~ 322

47 王啸波,唐玉秋,王金华,黄仪秀,陈润生,黄力. 环境样品中 DNA 的分离纯化和文库构建. 微生物学报, 2001,41:133 ~ 140

48 Loreng P, Liebeton K, Niehaus F, *et al*. Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. Current Opinion in Biotechnology, 2002, 13: 572 ~ 577

第4章　核酶和脱氧核酶定向分子进化

4.1　引　　言

RNA分子在现代生物学中有着重要的作用,特别是在最基本和高度保守的细胞过程。核糖体晶体结构揭示了RNA在蛋白质合成的作用[1~4]。它涉及作为DNA复制的引物,携带遗传信息到翻译装置,并且是位于核糖体“心脏”的催化剂。当拼接和编辑时,RNA指导前体mRNA的加工,并且调节细胞中一些其他RNA和蛋白质事件。核酶(ribozyme)协助RNA加工事件和病毒基因组的复制。个别的核苷酸作为重要的信号分子,并且它们的辅酶衍生物参与中心代谢的大多数反应。

尽管RNA只含有4种不同的碱基,但RNA可折叠成复杂的三级结构;与蛋白质类似,催化广泛的化学转化。实验室进行的RNA进化可以看作为RNA世界中RNA进化的一种模型,用于获得许多新的酶[5]。这些包括RNA催化核苷酸的合成[6]、RNA聚合[7]、tRNA的氨酰基化[8]和肽键的形成[9]。很可能RNA能够支持以RNA基因组为基础的生命,复制并通过RNA的催化功能来保持。然而,如何科学的了解RNA产生、代谢的复杂度、导致基因组和蛋白质组产生的方式,尚存在许多实质性缺口。

RNA是各种顺序的多聚物,通过模板机制自身复制。不同的顺序有不同的化学性质。但是几乎所有的顺序都能够形成Watson-Crick复体结构,以促进新拷贝的产生。如果RNA的构件在前生命环境是有效的,如果这些构件组合形成多核苷酸,如果某些多核苷酸开始自身复制,那么RNA可以被认为是地球上出现的第一个生命形式。但是根据目前生命化学的知识,未必是事实。核糖、磷酸、嘌呤和前嘧啶完全可能得到,虽然嘧啶有些勉强。它们可能以很低产率组合形成核苷,核苷酸及其类似物甚至可以形成多聚物(具有2′,5′-,3′,5′-和5′,5′-磷酸二酯键的混合物)、糖的D-立体异构体和L-立体异构体、糖苷键α-差向异构体和β-差向异构体和各种修饰的糖、磷酸和碱基。

了解依赖RNA的生命起源的主要障碍是确定合理的机制。几种研究途径或许有特殊的条件导致活化的β-D-核苷酸的优先合成,或这些单体优先移入多聚物中。例如,由甲醛合成糖导致核糖2, 4-二磷酸形成。当在某些金属氢氧化物存在下,这个反应可以在近中性pH水溶液发生。甲醛的聚合产生2′, 5′-键产物,而在白土(montmorillonite clay)存在下主要产生3′,5′-键产物。这样,通过一系列偏向合成,分级和其他的富集过程,形成RNA“汤”。

另一个途径是假设生命不是依赖 RNA 起始，而是依赖 RNA 之前的其他遗传系统。这个途径的研究最近几年有实质性进展。最值得注意的是 α-L-苏呋喃糖酰单元通过 3′,2′-磷酸二酯键相连形成的苏糖核酸（TNA）[图 4.1(a)]，这个类似物自身同 RNA 形成稳定的 Watson-Crick 配对，从克服前生命化学“混沌”的观点出发，TNA 较 RNA 有更多的优势，这是由于其化学性质简单。苏糖是仅有两个丁醛糖（aldotetrose）（四碳糖）的一种，并且只能在 2′和 3′位连接。不难想像“TNA 世界”可以过渡到“RNA 世界”。

另一个具有潜在的 RNA 的祖先是肽核酸（PNA）。PNA 是通过亚甲羰酰基把碱基连在 *N*-(2-氨乙基)甘氨酸单元的类肽骨架组成的[图 4.1(b)]。氨乙基甘氨酸是由氮、氨、甲烷和水在电火花反应中合成。PNA 自身或同 RNA 形成类 Watson-Crick 复体结构。PNA 单体可以经历分子内 *N*-乙酰转移反应以阻止它们按常规机制聚合。在 RNA 之前两种其他的可能是甘油衍生的核酸类似物[图 4.1(c)]和吡喃糖酰基 RNA（含 4′,2′-连接的 β-D-吡喃核糖酰基单元）[图 4.1(d)]，但两者都没有充分的实验支持。在依赖 RNA 生命之前也可能就有复制的进化多聚物，没有达尔文进化特征的自身复制的某些肽，甚至小的有机化合物[32]。TNA，PNA 或其他可能的 RNA 前体至今未见报道。

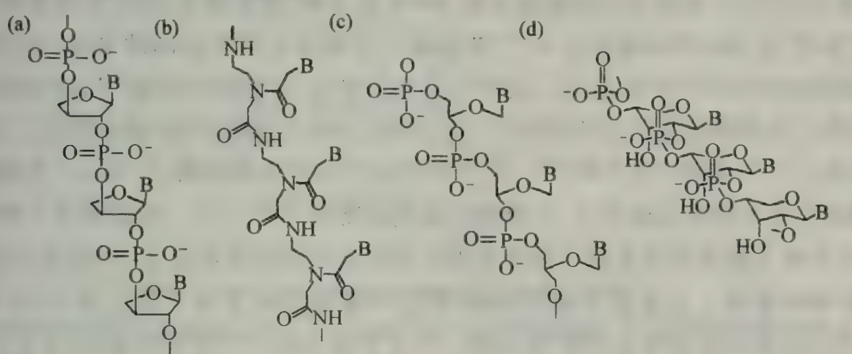


图 4.1　RNA 的可能前体

(a) 苏糖核酸（TNA）；(b) 肽核酸（PNA）；(c) 甘油衍生核酸（GNA）；
(d) 吡喃核糖酰基-RNA。图中 B 代表碱基

RNA 世界的核心过程是 RNA 的复制，且由 RNA 催化。复制可能涉及寡核苷酸或大的亚单位的连接，或者是通过模块组装，而不是线性模板的组织。有的定向进化的核酶是催化模板指导的寡核苷酸 3′-羟基和寡核苷酸 5′-三磷酸连接。最好的研究工作是类 1 连接酶（class 1 ligase）[10]，见图 4.2(a)。它含有一个内部的模板区，结合互补 RNA 底物，并且底物的 3′-羟基直接进攻核酶的 5′-三磷酸，形成 3′，5′-磷酸二酯键。约 120 个核苷酸组成的核酶催化速率接近 100/min，相对于未催

化反应,速率提高约 10^9 倍[11]。类 1 连接酶也催化 NTP 的聚合,按模板指导方式把 3 个残基加到 RNA 引物的 3′端[12][图 4.2(b)]。

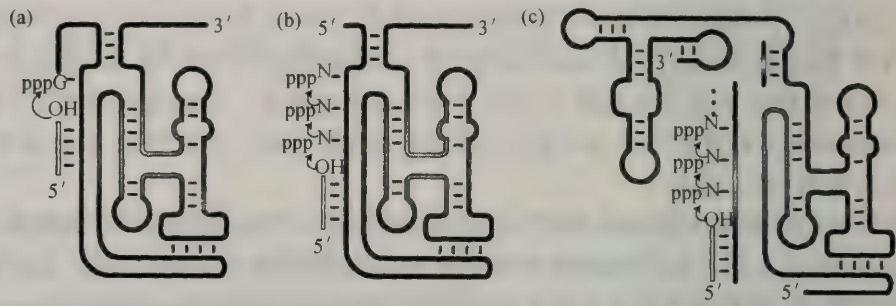


图 4.2 RNA 聚合酶的核酶体外进化的连续相

(a) 类 1 连接酶的核酶催化模板指导 RNA 引物(开线)的 3′端与核酶的 5′端连接;
(b) 类 1 连接酶通过内部模板指导催化 3 个核苷 5′三磷酸(NTP)加到引物的 3′端;
(c) 类 1 衍生的聚合酶催化 14 个 NTP 加到外部 RNA 模板上

类 1 连接酶作为进一步定向进化实验的起点,结果明显地提高 NTP 聚合的活性[图 4.2(c)]。最终进化的核酶含约 200 个核苷酸,并且催化 RNA 引物在外部模板上延伸,在 24h 内连续加上 14 个核苷酸。可能存在许多不同的 RNA 分子能够催化模板指导的 NTP 的聚合。由体外进化获得 hc 连接酶的核酶[13],在结构或顺序与类 1 连接酶没有明显的相似性,但也催化 RNA 引物在外部 RNA 模板上延伸,它只加上一个或两个核苷酸[14]。两个其他的定向进化的核酶,L1 和 L3 连接酶在内部而不是外部模板催化 3′,5′-磷酸二酯链的形成[15,16]。后一核酶只含 3 种核苷酸,完全缺乏胞嘧啶核苷酸。进化获得的其他连接酶还能催化 2′,5′-键而不是 3′,5′-键 RNA 形成。大量的可能的 RNA 顺序的探索可以预测产生一些 3′,5′-连接酶,其中许多可以进化为 NTP 聚合酶。对于含有 100 个核苷酸的 RNA 分子,就有 4^{100}(约 10^{60})种可能顺序。

通过体外进化已得到催化核苷酸合成的某些步骤的核酶(图 4.3),例如催化嘧啶和活化核糖形成核苷酸的核酶。在前生命化学中,这是个众所周知的难于进行的反应,但用随机顺序 RNA 库就可以实现该反应。随机顺序 RNA 结合在 5-磷酸核糖酰基-1-焦磷酸上,并允许同硫代尿嘧啶反应。进化酶催化反应的速率较之未催化反应提高 10^7 倍以上。另一个定向进化的核酶催化多核苷酸的 5′-磷酸化,应用 ATP 作为磷酸供体,催化速率提高约 10^9 倍[17]。催化模板指导的 RNA 连接的核酶已报道[18]。在核苷酸合成中有几个重要的反应至今未用核酶催化(图 4.3),例如由简单的醛形成核糖等。现在已证实通过体外进化所获得的核酶可以催化各种化学反应,包括酰基转移[19,20]、N-烷基化和 S-烷基化[21,22]、碳-碳键形

成[23,24]、酰胺键形成和 Michael 加成，但从未证实核酶催化核糖转为脱氧核糖[25,26]。

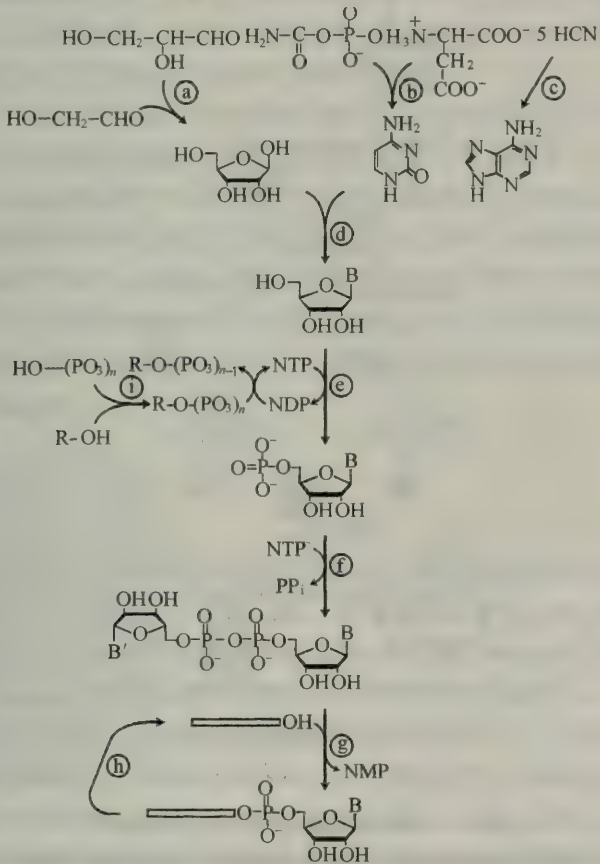


图 4.3　RNA 催化 RNA 合成的假说途径

带有圆圈的字母表示实验证实的反应。ⓐ 糖醛和甘油缩合形成核糖；ⓑ 氨基甲酰基磷酸的氨基甲酰基转移到天冬氨酸上，继之环化形成嘧啶；ⓒ HCN 的五聚合形成嘌呤；ⓓ 嘌呤或嘧啶(B)加到核糖上形成核苷；ⓔ 核苷磷酸化形成核苷酸；ⓕ 通过 NTP 的核苷酸部分转移核苷酸的活化；ⓖ 核苷酸加到 RNA 引物的 3′端(▭)，两个 RNA 底物被结合在互补模板的相邻位置(未标出)；ⓗ 核苷酸连续加入进一步延伸引物；ⓘ 由烷基磷酸将磷酰基转移到 NDP 产生 NTP，NMP 以相似方式转为 NDP。磷酸来源于聚磷酸矿

RNA 能够催化蛋白质合成的所有反应(图 4.4)。信息、转移和核糖体 RNA 分子存在于所有机体中。RNA 指导特殊多肽顺序的组装。在氨酰基腺苷酸形成中氨基酸被活化，继之氨基酸转移到 tRNA 的 2′(3′)端。在现代生物学中是通过一组 20 种氨酰基-tRNA 合成酶(蛋白质)催化的这些反应用定向进化的核酶也能完成。含有约 110 核苷酸的核酶催化亮氨酸或苯丙氨酸加到它特有的 5′端，形成氨酰基核苷酸酸酐[27]。有的核酶利用各种氨酰基-腺苷酸底物催化它特有的 2′(3′)端的

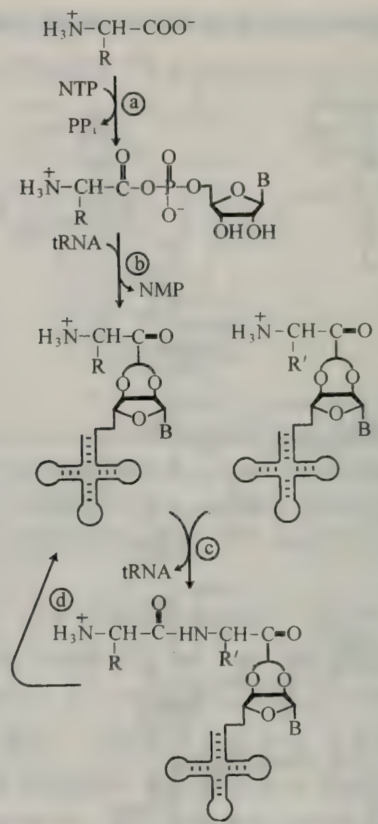


图 4.4　RNA 催化蛋白质合成的假设途径

圆圈字母表示反应已被实验证实。ⓐ 通过氨酰基核苷酸酸酐活化氨基酸；ⓑ 活化的氨基酸转移到 RNA 的 2′(3′)端，在 2′-氧和 3′-氧之间的半圆指出氨基酸在这两个位置迅速迁移；ⓒ 肽酰基转移结果形成二肽。两个氨酰基-tRNA 底物结合在互补模板的相邻位置(图中未标出)；ⓓ 连续的肽酰基转移反应形成多肽

氨酰基化[28,29]。另一核酶催化 tRNA 的氨酰基化，或共价系于核酶上或作为分离的底物[30]。蛋白质合成的最后步骤涉及沿 RNA 模板在相邻位置结合氨酰基和肽酰基的寡核苷酸，并通过氨基酸的 α 胺进攻肽酰基酯的羰基形成肽键。核糖体的晶体结构揭示肽酰基转移酶位点完全是由 RNA 组成的[31]。第一个核糖体较之现代核糖体是更小，更简单的结构，功能是有限的，很可能完全是由 RNA 组成的。E 位点是后来进化的。核糖体功能中心的化学组成基本上与它们的进化年代有关，越最近越含有更多的蛋白质。

定向进化也用于开发催化肽键合成的核酶。从 5′端带有苯丙氨酸的随机顺序 RNA 库起始，根据能够同蛋氨酰基-腺苷酸底物反应形成二肽进行选择。有一种

进化的核酶约含有 190 个核苷酸,并且催化速率约为 0.1/min。它接受几种不同的氨酰基腺苷酸,优选蛋氨酸。像类 1 连接酶的核酶进化为 RNA 聚合酶一样,这个肽酰基转移酶可以进化成连续合成多个肽键的核酶。

RNA 有几个特征使它适于作为简单达尔文体系的基础:它只含有 4 种不同的单元,当活化和结合到互补模板上时容易聚合;它是一种聚阴离子,不管顺序如何,易溶于水,并形成耐受顺序变动的简单二级结构[32,33]。这些特征同 DNA 和蛋白质比较,DNA 的较低的活性和更大的稳定性使其更好作为遗传材料,而蛋白质的化学多样性,包括阴离子、阳离子和疏水基团,使其更好的作为催化功能的基础。

在生物学上发现的核酶是少的,更多的是通过试管进化得到的。人工构建依赖 RNA 的生命是一个遥远将来的事件[34],"RNA 世界"的残迹可能潜在于某种特殊的现代微环境之中。

4.2 核　　酶

1981 年 Cech 等[35]发现四膜虫的前体 RNA 可以在没有蛋白质存在的情况下自身催化切除内含子,完成加工过程。这一具有催化活性的 RNA 的发现丰富和发展了"酶"的概念,从此对具有催化活性的 RNA,即核酶(ribozyme,RNAzyme,catalytic RNA,RNA enzyme)的结构、催化机制以及应用的研究日益深入。

目前为止,在自然界中发现的核酶根据其催化的反应可以分成两个大类:①剪切型核酶,这类核酶催化自身或者异体 RNA 的切割,相当于核酸内切酶,主要包括锤头型核酶,发夹型核酶,丁型肝炎病毒(HDV)核酶,以及有蛋白质参与协助完成催化的 RNaseP;②剪接型核酶,这类核酸主要包括组 I 内含子和组 II 内含子,实现 mRNA 前体自我拼接,具有核酸内切酶和连接酶两种活性。

4.2.1 锤头型核酶

由于锤头型核酶小,且分裂机制简单,所以它是最好的结构研究的候选者。Haseloff 和 Gerlach[36]提出锤头型核酶主要是由 3 个结构域组成[如图 4.5(a)]:①切割结构域,包含 GUX(X = A,U,C,≠G),以 GUC 作为切割点活性最高;②催化结构域,如图所示,其中 13 个核苷酸是绝对保守的,它是酶活性所必需的序列。这 13 个核苷酸是 5′GAAC(N)*n*CUX(N)*n*CUGA(N)*n*GA3′;③结合结构域,即核酶两侧与靶 RNA 碱基配对区域。研究表明,配对区长度为 12 碱基的反应具有更高的 K_{cat}/K_m 值。双螺旋 I 短于双螺旋 III 的核酶比双螺旋 I 长于双螺旋 III 的核酶切割速率大 1～2 数量级,当双螺旋 III 长度为 10bp 时,双螺旋 I 的最佳长度为 5bp[37]。

核酶属于金属酶，金属离子为核酶辅因子，需要二价金属离子，例如 Mg^{2+}。核酶的活性形式是一种 RNA 键合金属氢氧化物。

所有锤头核酶呈球状构象，可以粗略地看作 γ 型折叠[图 4.5(b)][38]。螺旋 Ⅱ 和螺旋 Ⅲ 几乎共轴，螺旋 Ⅱ 由两个 GA 增长，逆向 Hoogsteen 碱基配对和一个不寻常的 AU 碱基配对，直接堆积在螺旋 Ⅲ 上，形成一个伪连续螺旋[见图 4.5(b)]。为了便于应用，一些科学家将锤头型核酶[图 4.5(a)]中茎-突环 Ⅱ 删除[图 4.5(c)]成为 Minizyme[39]。动力学和 NMR 分析指出其单体无活性，而二聚体表现出相当高的分裂活性。此外，两个不同的单体构成异源二聚体，第二个底物的存在促进二聚体的形成，并且分裂活性提高，活性与第二个底物的浓度成正比。这说明 Minizyme 其功能单位是二聚体 Maxizyme，其有两个底物结合区，识别顺序是不同的。它可同时作用在单一底物 mRNA 上的两个不同的靶位。这样异源二聚体

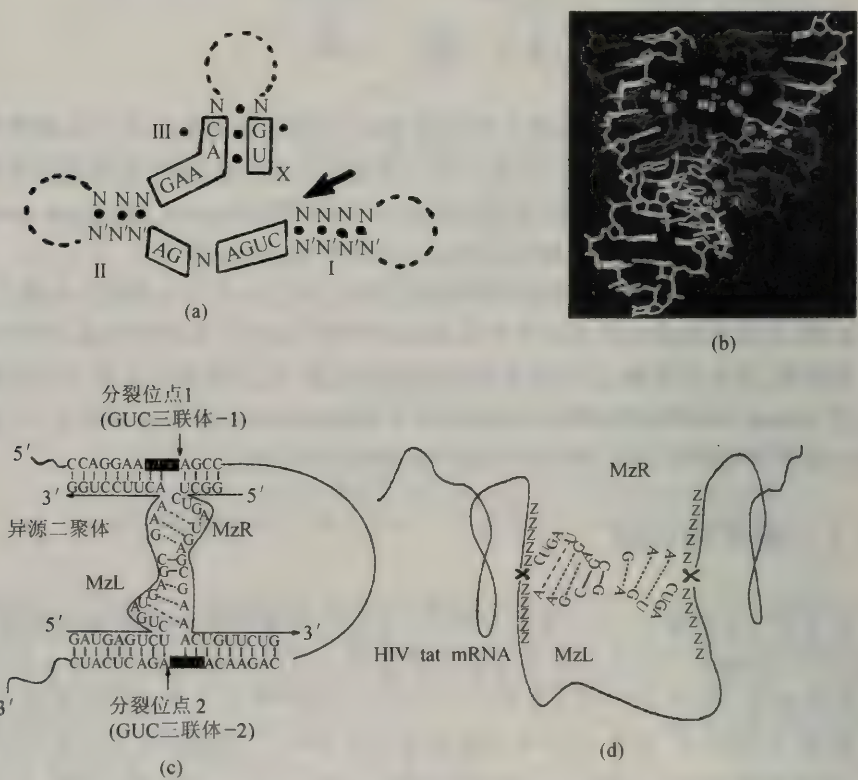


图 4.5 锤头型核酶的二级结构和空间立体结构示意图

(a) N,N′代表任意核苷酸；X 可以是 A、U 或者 C，但不能是 G；Ⅰ、Ⅱ、Ⅲ 是锤头结构中的双螺旋区；箭头指向切割位点。(b) 锤头型核酶的立体结构模型，白色链是核酶，灰色链是底物 RNA 分子，在磷酸骨架上结合有镁离子。(c) 异源二聚体 Maxizyme 的二级结构。(d) Maxizyme 切割 HIV tat mRNA

Maxizyme 具有高活性,并在体内外成功地分裂 HIV-1 tat mRNA[图 4.5(d)]。Maxizyme 可以有效地用于失活有害基因的表达。

Lott 等[40,41]提出了锤头型核酶催化反应的两种可能的化学机制:"单金属氢氧化物离子模型"(one-metal-hydroxide-ion)机制[图 4.6(a)]和"双金属氢氧化物离子模型"[图 4.6(b)]。图 4.6(a)中金属氢氧化物作为广义碱从 2′-羟基获得一个质子,这个被活化了的 2′-羟基作为亲核基团攻击切割位点的磷酸。图 4.6(b)中 A 位点的金属离子作为 Lewis 酸接收 2′羟基的电子,其极化并减弱了 O-H 键,使 2′-羟基中的质子更容易离去。B 位点的金属离子也作为 Lewis 酸接收 5′羟基的电子,极化并减弱了 O-P 键,使 O 成为更容易离去的基团。HDV 中的核酶显示了与以上不同的催化机制[图 4.2(c)],胞嘧啶(C^{76})充当一般碱,咪唑环上的 N 吸引 2′

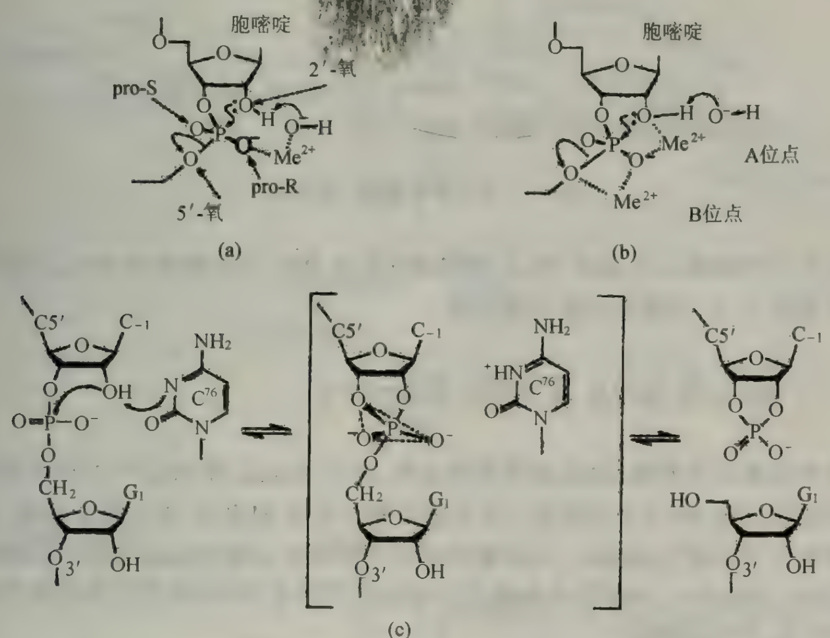


图 4.6　锤头型核酶的两种可能的催化机制以及 HDV 核酶的催化机制
(a)单金属氢氧化物离子模型;(b)双金属离子模型;(c)HDV 核酶中胞嘧啶充当一般碱进行催化的反应机制

羟基上的质子,活化了羟基上的 O,这个 O 亲核攻击相邻的核酸骨架上的 P,经过过渡态形成磷酸内酯键,而原来核酸骨架的磷酸酯键断裂。目前的研究还没有证据显示包括金属离子在内的其他活性基团如何发挥作用。

4.2.2 发夹型核酶

图 4.7 是一个发夹型核酶[42]的二级结构模型。50 碱基的核酶和 14 碱基的底物形成了发夹状的二级结构,包括 4 个螺旋和 5 个突环。螺旋 3 和 4 在核酶内部形成,螺旋 1(6 碱基对)和 2(4 碱基对)由核酶与底物共同形成,实现了酶与底物的结合。核酶的识别顺序是(G/C/U)NGUC,其中 N 代表任何一种核苷酸,这个顺序位于螺旋 1 和 2 之间的底物 RNA 链上,切割反应发生在 N 和 G 之间。

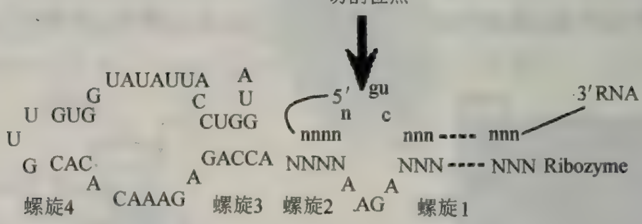


图 4.7 发夹型核酶的二级结构

以上两种核酶以及 HDV RNA,链孢霉线粒体 RNA 等切割底物 RNA 后的产物都是 3′端的 2′,3′-环磷键及新 5′端羟基。

4.2.3 蛋白质-RNA 复合酶(RNaseP)

这类核酶主要催化 tRNA 前体成熟过程,例如 Altman 和 Pace[43,44]两个研究组发现的大肠杆菌 tRNA 5′成熟酶。这个酶由蛋白质和 M_1RNA 两个组分构成,其中蛋白质的分子质量为 20kDa。实验证明来自不同原核细胞 RNaseP 中的 M_1RNA 具有相似的三维结构。与前面几种剪切型核酶不同的是,RNaseP 催化得到的产物的 3′端是羟基,5′端是磷酸。

E. coli RNaseP 以顺式分裂前体 tRNA 底物。通过缺失 tRNA 的 5′端片段构建了顺序特异的内切核酸酶(图 4.8),通过把指导顺序连接到 M_1RNA 的 3′端,设计出顺序特异的核酶(图 4.9)。

两种类型 RNaseP 核酶:带有内部指导顺序的环状置换的 *E. coli* RNaseP RNA 称为 EndoP;指导顺序共价系在 RNaseP RNA 的 3′端的 *E. coli* RNaseP RNA,称之 M_1GSRNA。

E. coli RNaseP 中的 M_1RNA 识别目标主要根据它的结构,这样可以在不需要目标 mRNA 的特殊顺序的情况下,设计核酶以破坏其他分子。M_1RNA 是消除染色

体移位后肿瘤发生的理想工具[45]。

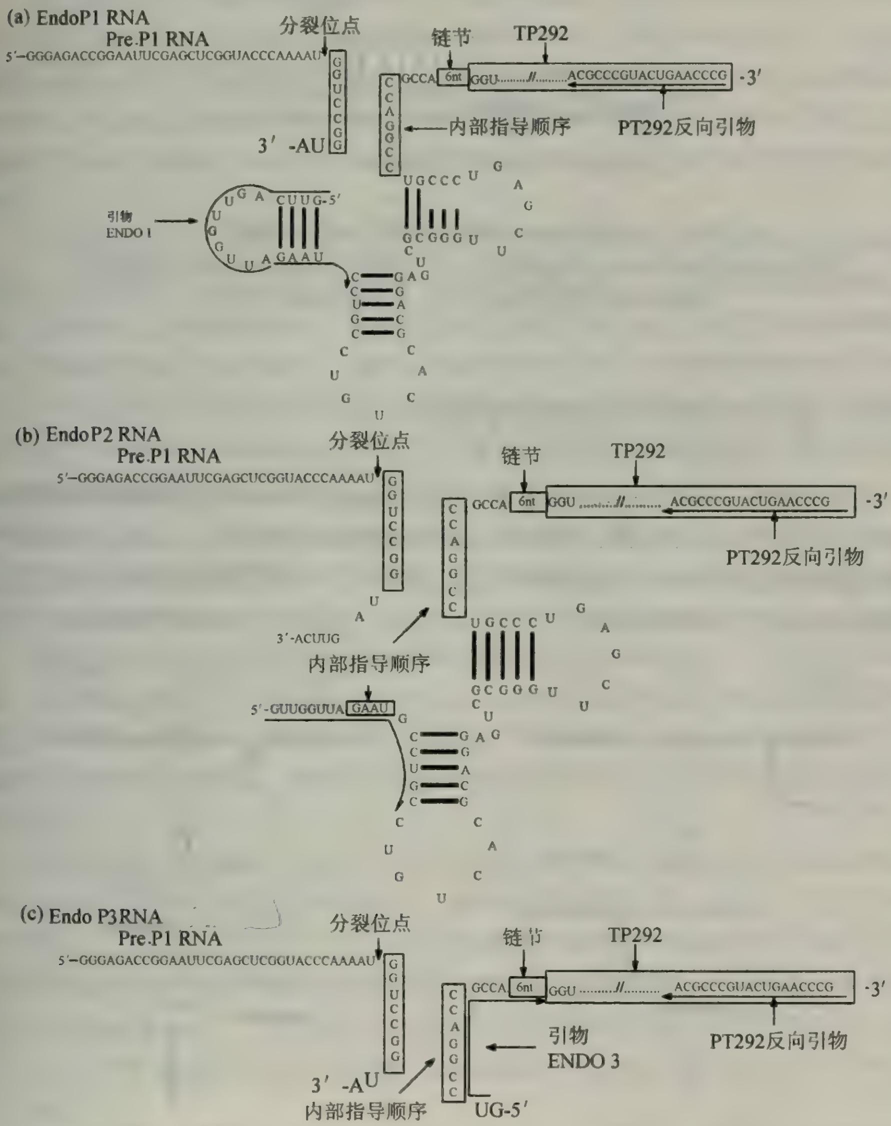


图 4.8　根据 RNaseP 核酶，设计特异顺序 EndoP 核酶

箭头指出在外源底物 RNA 内的分裂位点；底下划线顺序代表 PCR 正向和反向引物的位置；EndoP3 RNA 是理论上的核酶，实验并未证实

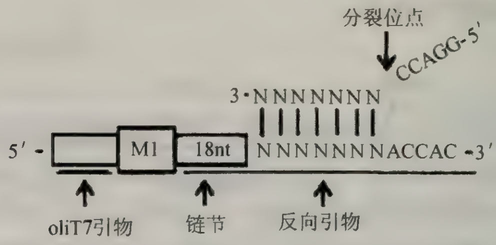


图 4.9　根据 RNaseP 核酶，特异顺序 MIGS 核酶的设计
箭头表示在外源底物 RNA 内的分裂位点。底下划线顺序代表 PCR 正向和反向引物。
内部指导顺序用核苷酸(N)表示，并与分裂的靶顺序互补

4.2.4　组 I 内含子和组 II 内含子

组 I 内含子(group I intron)和组 II 内含子(group II intron)，这类核酶比较复杂，通常包括 200 个以上核苷酸，主要催化 mRNA 前体的拼接反应。Cech 及其同事[46]发现四膜虫的前体 RNA 可以在体外无蛋白质参与下切掉它自身 413nt 内含子。这就是由组 I 内含子核酶催化的反应[图 4.10(a)]，包括两个连续的转酯反应，并且

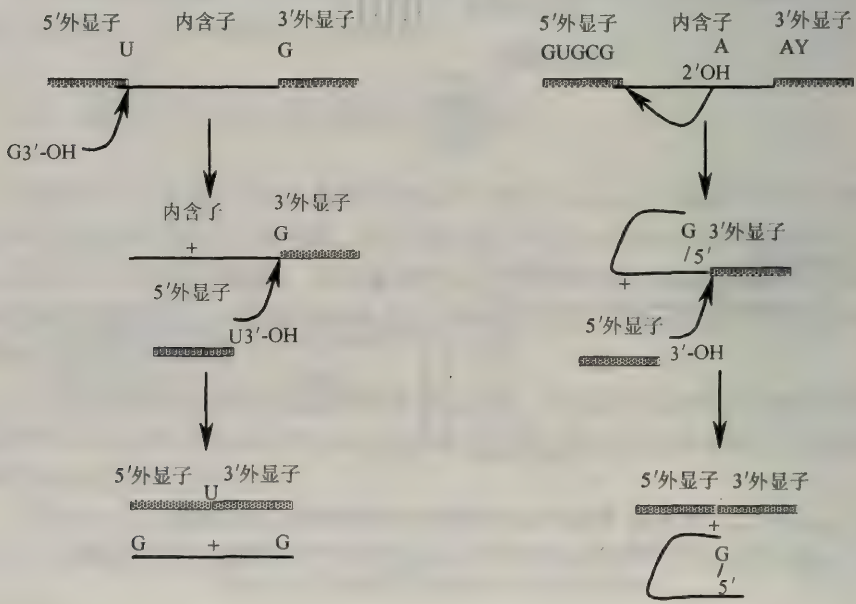


图 4.10　组 I 内含子(a)和组 II 内含子(b)的剪接机制

需要 Mg^{2+} 或 Mn^{2+} 及鸟苷(或鸟苷酸)的参与。

组 I 内含子的界限可以简单地用 5′外显子(exon)3′端的 U 和内含子 3′端的 G 来界定。组 I 内含子能够自身剪接是与它们保守的二级和三级结构有关。像蛋白质酶一样,内含子形成高级结构的折叠结果使关键残基形成活性部位,在辅助因子的参与下实现自身剪接。除了剪接之外,组 I 内含子还可催化各种分子间反应,包括剪切 RNA 和 DNA、RNA 聚合、核苷酰转移、模板 RNA 连接以及氨酰基酯解等[47]。

与组 I 内含子一样,在体外,组 II 内含子的剪接是经过两个转酯化反应来实现的,无蛋白质参与。组 I 和组 II 内含子的主要差别是第一步反应的化学机制。在组 I 内含子中,外部的鸟苷的 3′羟基作为进攻基团,而在组 II 内含子中是内部腺苷的 2′羟基起作用[图 4.10(b)]。这个反应的结果形成一个带突环的内含子 3′外显子分子,其中第一个核苷酸经由 2′,5′-磷酸二酯键与内含子的 A 相连。在第二步反应中,5′外显子的 3′羟基进攻内含子 3′外显子连接点,结果是两个外显子相连,并释放出带有突环的内含子。

4.2.5　环状核酶[48]

利用 RNA 拼接反应的方法在体外和体内产生环状 RNA 是一个正在发展的领域。这些方法将促进环状 RNA 的研究,并且最终将实现目标 RNA 在体内以环状表达。环状 RNA 除了它们在生物化学研究中的价值外,在基因表达调节和治疗领域都有潜在的应用价值。然而天然环状 RNA 是少见的,它们在体内的产生或消亡所知甚少。

体外制备环状 RNA 的方法主要有两种:化学法和酶法。它们都需要线性 RNA 的合成和纯化,继之成环并纯化。自我拼接内含子(组 I 或组 II)已用于外显子的环化。在体外拼接反应中外显子顺序促进环状产生,并且可以提供体内目标顺序环化的一般方法。这种方法尤其适于特殊细胞类型或组织的剪接小体内含子(spliceosomal intron)。三股环状 RNA 已经由线性 RNA 制成,调节拼接反应的核酶可以用于产生环状的反式作用核酶(*trans*-acting ribozyme),环状 RNA 可以作为反义 RNA。

环状 RNA 具有各种优点:①环状较之线性 RNA 更稳定,例如发夹结构的环化可以阻止分子间二聚体的形成,环状 RNA 抗外切酶的降酶;②体内半衰期长。部分实验表明 RNA 末端的性质贡献给稳定性。一些资料指出,在细胞中和细胞提取液中,环状 RNA 较之相同或相似顺序的线性形式更稳定。

天然发生的环状 RNA 最著名的例子有植物和酵母中小的环状 RNA 复制子。环状结构最基本的功能是提供滚动环复制的模板。

在植物中,病毒卫星 RNA 和类病毒 RNA 是以环状形式存在,RNA 是复制环的一部分,其含有自我分裂的顺序(核酶),它可以精确地分裂复制时产生的 RNA 连环体。

在烟草环斑病毒中发现的发夹核酶,可以在体外有效地催化正常分裂反应中逆向环化,这表明同一核酶可以兼有连环体的分裂和线性单体的环化。在体内由转录表达的发夹核酶成环的方法也已描述,由于反应的可逆性,环状与线性处于平衡之中。锤头核酶分裂反应中逆向调节不如发夹核酶,但是通过细胞中的酶催化连接形成环状仍是可能的。

肝炎 δ 病毒(HDV)(一种人类的病原),与某些植物致病 RNA 有共性。HDV 含有环状 RNA 基因组并在基因组和反基因组中含自身分裂的核酶。HDV 基因组大于植物 RNA 病原,大约 1700 个核苷酸,然而它小于哺乳动物病毒的基因组。应用 HDV 作为载体导入特殊的 RNA 顺序,例如反式作用核酶是可行的。RNA 基因组大概通过滚动环机制来复制。HDV 核酶在结构上不同于锤头和发夹核酶。逆反应(连接)在体外没有证实,某些证据指出单体的环化可能是由宿主催化完成的。

在 *Saccharomyces cervisiae* 中检出小的单股 RNA,当在醋酸钾介质中生长时,某些菌株积累了 20S RNA,一些证据表明这个单股 RNA 可以环状形式积累。与植物和 HDV 不同,酵母 20S RNA 可以为 RNA 聚合酶编码。

在拼接反应中环状外显子的产生首先是用组 I 内含子证实的,它可以提供在体内外产生各种环状 RNA 有效途径。

(1) 组 I 内含子

修饰组 I 内含子在产生环状外显子似乎是很有效的。在 3′外显子下游 5′拼接位点的活化结果在拼接的 3′外显子产生环状形式。四膜虫组 I 内含子-外显子(PIE)顺序有效地拼接产生环状外显子,含有 3′半内含子、外显子和 5′半内含子[图 4.11(a)]。

(2) 组 II 内含子和 Pre-mRNA 剪接小体

如果顺序被重排的话,组 II 自身拼接顺序也产生环状外显子[图 4.11(b)]。这个反应似乎弱于组 I 的 PIE 顺序,但或许它在体内更有效。

Pre-mRNA 拼接装置可以开拓成产生环状 RNA 的装置。剪接小体内含子的拼接和组 II 内含子相似。

拼接方法产生环状 RNA 有许多优点:①在体外产生环状的反应简单。这个反应犹如转录。在某些情况下仅涉及增加 Mg^{2+} 浓度,提高温度,或者两者兼而有之;②环状 RNA 容易纯化;③具有在体内制备环状的潜力。它可以表达且稳定的具有治疗价值的 RNA 顺序。

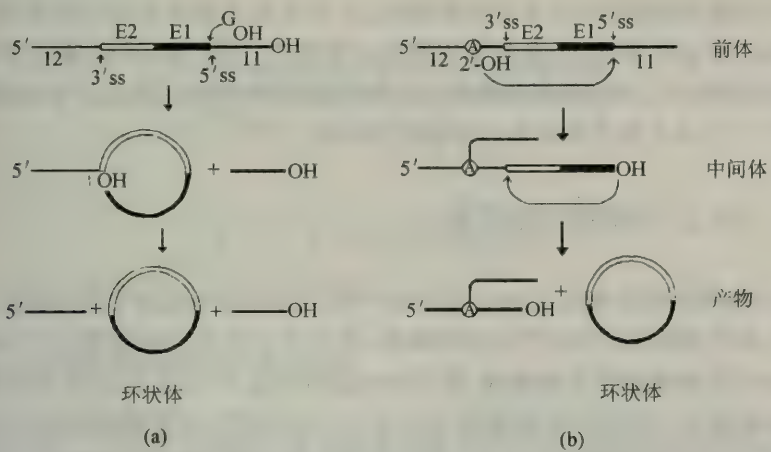


图 4.11　A、B 由 PIE 顺序形成环状的机制

(a) 组 I PIE 顺序的拼接;(b) 组 II PIE 顺序的拼接。11 内含子的 5′部分;12 内含子的 3′部分;E2 和 E1 融合的外显子顺序

4.3　脱氧核酶

脱氧核酶(deoxyribozyme or DNAzyme or DNA enzyme)或称催化 DNA(catalytic DNA)是 20 世纪 90 年代初才被人们所认识。最初的设想是来自 RNA 酶中部分核苷酸被相应的脱氧核苷酸所代替后仍具有酶活性,那么在理论上应该能得到具有酶活性的 DNA 分子。1994 年 Breaker 等[49]通过体外筛选方法首先得到了依赖于 Pb^{2+} 具有切割 RNA 分子功能的 DNA 分子,并将其命名为脱氧核酶。

虽然 DNA 分子在自然界中几乎都是以双链形式存在,但特殊条件下以单链形式存在的 DNA 分子在理论上应该可能出现与人工合成的脱氧核酶相似的结构,但迄今为止尚未发现天然存在的脱氧核酶,目前报道的脱氧核酶都是通过体外选择/筛选得到的。由于脱氧核酶具有分子相对较小、结构稳定、催化效率高、易于合成和修饰等优点,在基因治疗、疾病诊断及分子生物学研究等方面都有诱人的应用前景。

DNA 分子由于没有 2′羟基,分子刚性增加,通常被视为不活泼分子,适于携带、传递遗传信息,在生物体内常以双链形式存在。而脱氧核酶都是单链 DNA 分子,可以通过自身卷曲、折叠形成特殊的三维结构,在某些特殊的辅助因子作用下,与底物结合并发挥其催化功能。

脱氧核酶通常分子较小(目前报道最小的脱氧核酶仅 25nt,其活性中心

11nt)[50],与蛋白质酶相比分子多样性较差,人们也在考虑将其他一些活性基团引入其结构。比如将氨基、咪唑基等活性基团修饰的 dATP 或 dUTP 通过 PCR 方法引入脱氧核酶分子,增加脱氧核酶分子多样性,以期得到高活性或具有新功能的脱氧核酶[51]。下面介绍几种典型的脱氧核酶结构。

4.3.1 10-23 结构脱氧核酶

10-23 结构脱氧核酶是 Santoto 和 Joyce 等[52]通过体外选择方法得到的。该酶相当于一种序列特异性限制性内切核酸酶,活性中心区为"10-23 型主"(10-23 motif),由 15 个脱氧核糖核苷酸组成,第八个碱基可以是 T、C 或 A,以 T 活性最高,其余序列高度保守。活性中心两端各为一长 7 ~ 9nt 的臂,为底物结合部位,可通过碱基互补与底物 RNA 特异性结合,其序列可以根据靶 RNA 的改变而灵活变动。10-23 结构脱氧核酶的靶部位为 5′…R↓Y…3′(R = A 或 G, Y = U 或 C)之间的磷酸二酯键,其中 R 不形成碱基配对,Y 则必须与脱氧核酶形成碱基配对。5′…A↓U…3′和 5′…G↓U…3′相对易被切开[52,53](图 4.12)。

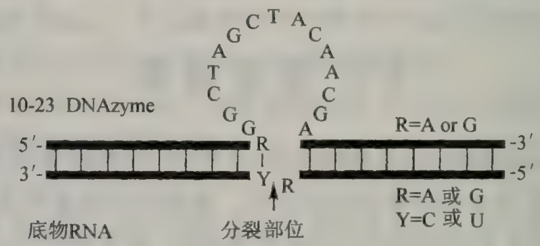


图 4.12 10-23 型主脱氧核酶的结构

4.3.2 8-17 结构脱氧核酶

这种结构类似锤头核酶且具有 RNA 切割活性的脱氧核酶,它也是 Santoto 等通过体外选择方法得到的,其活性中心为 8-17 型主(8-17 motif),包含一个常为 3 个碱基对的茎-环结构和一个 4 ~ 5nt 的非配对区。茎至少应为 2 个 G-C 对。茎环 5′-AGC-3′高度保守,非配对区一般为 5′-WCGR-3′或 5′-WCGAA-3′(W = A 或 T, R = A 或 G)。切割位点是 5′-A↓G-3′, A 不与脱氧核苷酸配对, G 则必须与脱氧核苷酸配对。Li 等[53]通过体外选择得到与 8-17 结构类似但依赖 Zn^{2+} 结构的脱氧核酶,该酶可切割 RNA 分子或 RNA/DNA 杂合分子,切割位点为 rNG(rN 为核糖核苷酸, G 为核糖鸟苷酸或脱氧核糖鸟苷酸)[54]。8-17 型主脱氧核酶结构见图 4.13。

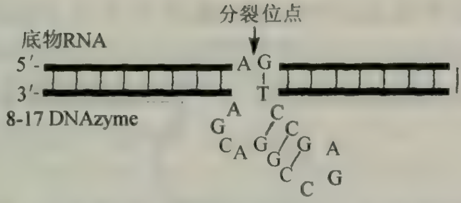


图 4.13　8-17 型主脱氧核酶的结构

4.3.3　手枪型结构脱氧核酶

Carmi 等[55,56]报道了两种二级结构呈手枪状具有自切割功能的 DNA 分子。I 型(class I)为 69nt 单链 DNA 分子,分子内形成 3 个碱基配对区,其自身切割活性依赖于 Cu^{2+} 和抗坏血酸盐。II 型(class II)为 46nt,分子内有两个碱基配对区,只依赖于 Cu^{2+}。自身切割位点在第 14nt 处(图 4.14),其 3′端约 27 个碱基对自身切割活性的发挥至关重要。若将 C_{35} G_{43}中的 C_{35}换为 G,或将 G_{43}换为 C,均使活性显著降低;但若将此配对位置对调为 G_{35} C_{43},则仍能保持一部分活性;第 1 ~ 7nt 发生突变或被删除,其活性显著下降。

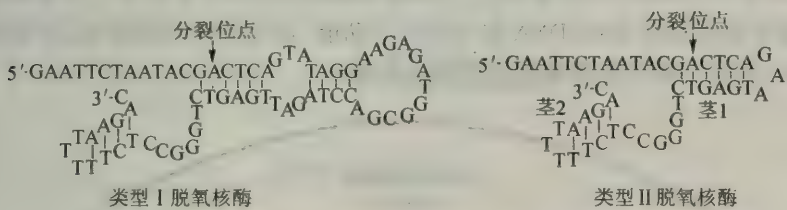


图 4.14　分裂 DNA 的手枪型脱氧核酶(类型 I、类型 II)

4.3.4　"二分"型结构脱氧核酶

"二分"型结构脱氧核酶(bipartite DNAzyme)是一类比较特殊的脱氧核酶,其活性中心 20nt,底物结合区 5 ~ 8nt,具有 RNA 切割活性。识别底物 RNA 位点为 A↑ANNN(N 为任何一种核苷酸,不同组合切割活性有一定差异)[57],其通用结构见图 4.15。

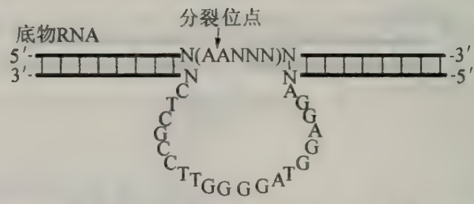


图 4.15 分裂 RNA 的“二分”型结构脱氧核酶

4.3.5 环状结构脱氧核酶

目前，脱氧核酶在应用上仍存在诸如在生物体内稳定性及如何有效地导入生物体到达靶部位等问题。我们将 10-23 结构脱氧核酶分子引入一环状结构，一方面通过环状结构及较大分子结构来提高其稳定性，一方面可使脱氧核酶能在生物体内遗传。

我们以 β-内酰胺酶 mRNA 为靶位点，设计并合成 10-23 结构脱氧核酶，将其克隆到噬菌体 M13mp18 质粒中，通过噬菌体表达系统产生含有脱氧核酶片段的单链环状 DNA 分子(DzM13)(图 4.16)。体外实验表明，DzM13 在某些二价金属离子辅助下具有特异性切割 RNA-DNA 嵌合底物的功能，催化特性与报道 10-23 结构脱氧核酶相近，提示二者可能遵循类似的催化机制。

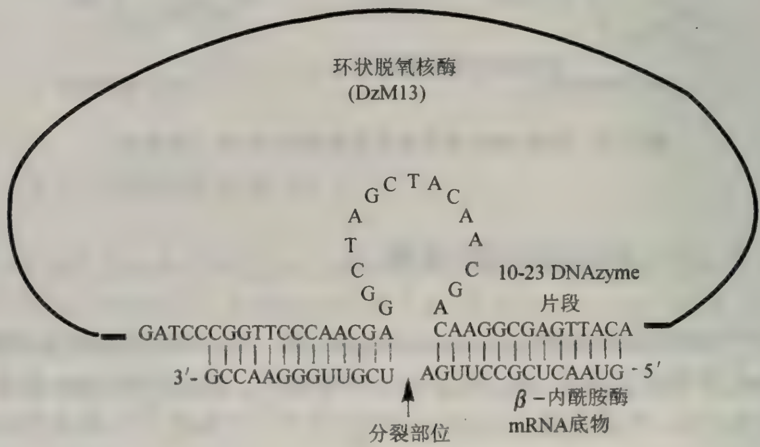


图 4.16 单股环状脱氧核酶(DzM13)的结构

实验结果表明：DzM13 对 S1 核酸酶的稳定性远远高于线状或硫代修饰的脱氧核酶。该环状结构不仅使脱氧核酶稳定性明显提高，更值得注意的是所构建的 DzM13 能在宿主体内复制并能连续传代。这种构建脱氧核酶的策略对脱氧核酶在实际中的应用及低成本生产都具有一定意义。

4.4　核酶和脱氧核酶的体外进化

4.4.1　拓展核酸顺序空间的新方法

过去 20 年，体外进化已经产生了能够催化各种化学反应的 DNA 和 RNA 分子(图 4.17)。体外进化 DNA 和 RNA 新方法的开发拓展了核酸的顺序空间，并扩大了具有化学功能的核酸库，将进一步增加产生功能分子的途径。

体外进化可以产生最低生命形式所需要的核酸催化剂。从以前进化的RNA

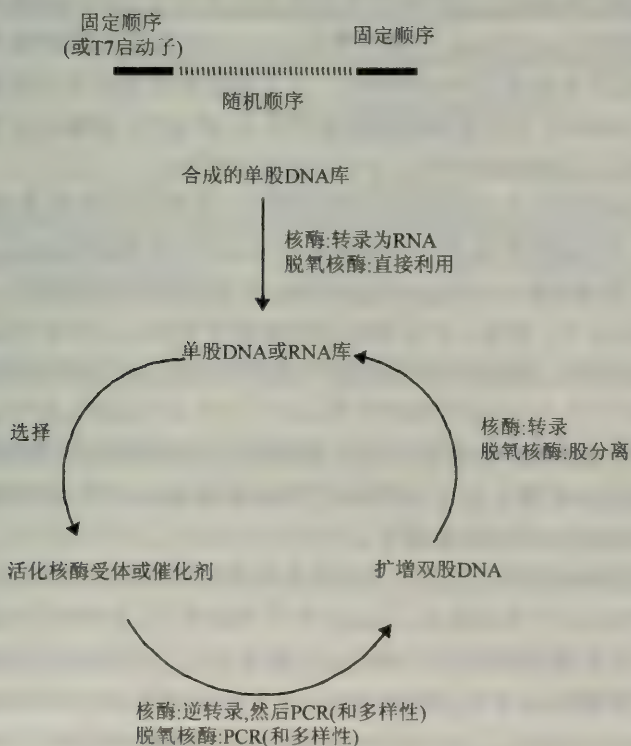


图 4.17　核酸的定向进化

随机合成的 DNA 库或者根据结合或催化活性直接选择，或者在选择前转录成 RNA

连接酶的核酶出发,Bittker 等[61]进化一种依赖 RNA 的 RNA 聚合酶,这是"RNA 世界"的一种关键酶。当用 RNA 引物-模板复合物和等摩尔浓度的 NTP 时,进化的核酶被证实有 97%顺序忠实性[62],24h 能够延伸 11 个碱基。

体外进化 RNA 和 DNA 分子促进了与前生命世界相关的其他反应的发现,包括嘧啶磷酸核糖化[63],寡核苷酸腺苷化[64]。虽然这些研究证实核酸是能够催化原始生命主要的反应,但不知道体外进化组装的生命是否在几亿年前就存在,或第一个催化剂重新组装成现代核酸结构[65]。

从核酶和脱氧核酶体外进化可以看出天然发生的 RNA 催化剂如何进化。Salehi-Ashtiani 和 Szostak 等[66]由随机 RNA 库分离出一些分裂 RNA 的核酶，发现与天然发生的锤头核酶型主惊人的相似，这个发现指出趋同进化途径可以产生在自然界所观察到的有限量的自我分裂的核酶。体外进化核酶也用于证实新 RNA 折叠可以由不相关的折叠进化，不需要无活性的中间体[67]。这些结果指出既不是功能也不是结构相似的核酸也可能是由共同的祖先进化来的。

进化核酸受体和催化剂的新方法的开发将提高它们的应用潜力和扩展应用范围。例如代谢组(metabolome)或蛋白质组(proteome)检验可能需要大量的反复选择和扩增循环以产生组成适体(aptamer)，这个过程的自动化可以加速体外进化，原型核酸进化自动机，可以在两天内完成 8 种 12 轮选择。能够结合鸡卵清溶菌酶的 RNA 适体进化，从初始随机库出发大约富集 10^{12}倍。

现有的核酸体外进化的方法,包括SELEX(选择和扩增重复循环而没有在各轮之间产生附加的突变)[68],盒式突变SELEX,易错PCR SELEX,每轮PCR引入的点突变频率1%～10%。SELEX不是探索核酸的顺序空间,而是从起始顺序库富集活性分子。相反，易错 PCR 是探索顺序空间。点突变可以提高所要求的结合或催化性质，提高点突变速率可以在顺序空间中接近高活性的突变体(图 4.18)。

除了开发核酸顺序空间新方法外,核酸进化的范围也随着附加功能团的 DNA 或 RNA 的进化而扩展。这个途径取决于 DNA 和 RNA 聚合酶是否能够接受修饰核苷酸,这方面的工作已有报道[61]。

结合凝血酶的适体是由含 5-戊基脱氧尿嘧啶核苷 DNA 随机库进化的,这表明修饰核苷酸可以用于体外进化。Eaton 等[69]用尿苷 C5 修饰的 RNA 库选择获得第一个催化 C-C 形成的催化剂(Diels-Alder 催化剂)。这个进化的 RNA 的催化活性取决于 Cu^{2+} 和修饰碱基的存在,暗示 Cu^{2+} 与吡啶基部位配位可能在 RNA 结构或者在指导 Diels-Alder 反应起关键作用。应用两个修饰的核苷酸同时进化的例子已报道。Perrin 等[70]成功地进化了 RNase 模拟物,其包含咪唑修饰的脱氧腺苷和氨基修饰的脱氧尿苷,这样将分裂 RNA 的蛋白质酶活性部位的两个功能团引入核酸;另一个依赖 DNA 的 RNase 模拟物(含有咪唑功能团的脱氧尿苷),在 Zn^{2+} 存在

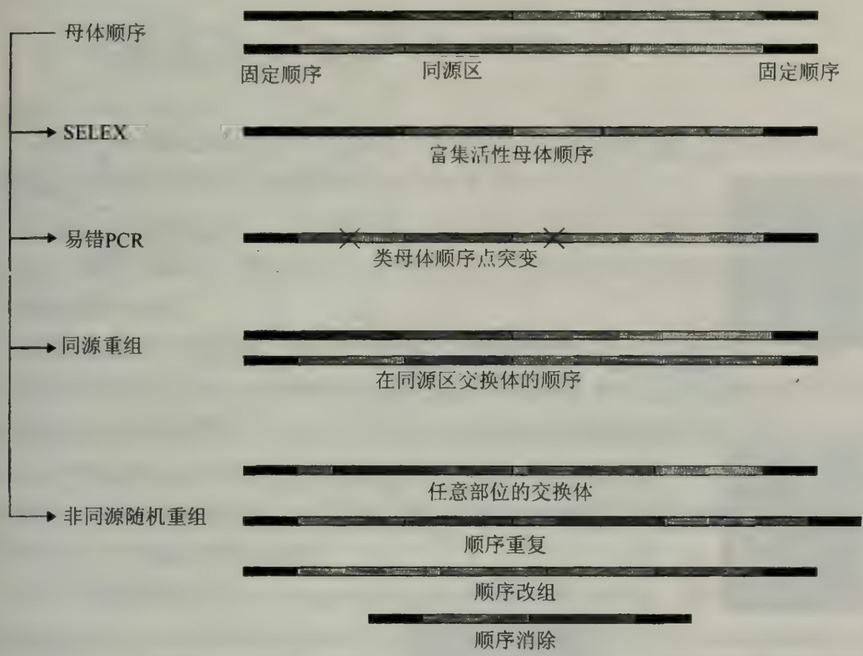


图4.18　核酸进化的多样性方法

以母体顺序起始，用SELEX富集活性母体顺序；用易错PCR使母体顺序产生点突变；同源重组，例如DNA改组，允许交换体只在同源区域之间产生；非同源随机重组，包括任何顺序之间的随机交换，顺序重复，改组和消失以及任意重组

下进化的模拟物分裂RNA。至今，研究人员应用聚合酶已把20种以上功能修饰核苷酸移入DNA或RNA(图4.19)。DNA聚合酶能够耐受C5广泛修饰的尿苷三磷酸。C5修饰的脱氧尿苷链节可以调节聚合酶效率。Williams等[71]研究表明Taq DNA聚合酶很好地耐受7-脱氮杂脱氧腺苷类似物，可以未修饰的dATP相似的效率移入DNA，这个结果指出7位对结构修饰有高度耐受性。Famulok等[72]同时完成了4个修饰的dNTP移入79-聚体DNA中，结果形成的修饰DNA股成功地转录和顺序化。

核苷酸修饰已集中在尿苷的C5[73]。应用T7RNA聚合酶可以成功地将这些修饰核苷酸移入RNA中，但比天然NTP效率低。研究人员证实修饰核酸所催化的反应不能用未修饰核酸代替，这样扩大了功能化核苷酸在核酸进化中的应用。

目前，核酸功能化已集注在将类蛋白质功能团移入寡核苷酸中。例如，应用聚合酶将咪唑、酚和一级胺作为组氨酸，酪氨酸和赖氨酸的代理移入DNA或

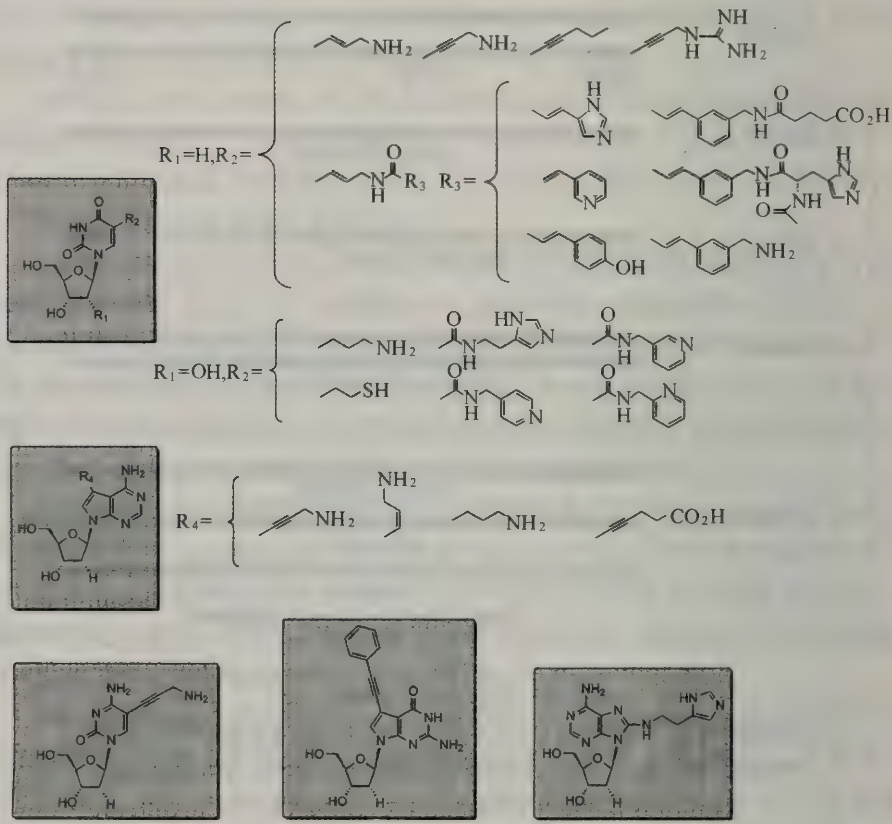


图 4.19 修饰核苷酸作为 DNA 或 RNA 聚合酶的底物

RNA。核酸同蛋白质比较，多样性和构象是有限的。然而，试图将类蛋白基团加到核酸碱基上也是可能的，尤其试图用修饰核酸模拟天然蛋白质。至少可以在聚合酶耐受的位置将非天然功能团引入核酸库。相反，如果把非天然功能团移入蛋白质中将会破坏蛋白质生物合成机制[74~78]。这说明功能团移入核酸中则是可行的，因此有很大的前景。在某种情况下虽然核酸不能与蛋白质催化潜力相匹敌，把 adhoc 功能移入 DNA 或 RNA 库可以使核酸能够解决蛋白质难以解决的问题。例如，用结合金属的配基(例如膦类)可以形成新的过渡金属配位复合物，其可以对成键或断键的催化进行定向选择，继之 PCR 扩增和多样化。此外，核酸的构象刚性和缺乏疏水部位限制其结合小分子。通过柔性链节把疏水基团移入核酸碱基上，可以增加它们结合小分子的能力，并具有高亲和力和特异性。指导体外进化的动力是选择，而不是筛选。 10^{15}金属和小分子结合的核酸库，能够立体选择性

的催化自然界没发现的化学反应，这是该领域诱人的可能性之一。

4.4.2　体外选择别构核酶

一般工程别构核酶或脱氧核酶有三种策略：模块合理设计；体外选择和组合设计（模块合理设计和体外选择的组合）[79,80]。

典型的别构酶包括三个功能特征：一个催化中心，一个效应因子结合中心和一个结构上敏感的结构。当效应因子结合时允许催化中心功能的改变，或者活化或者抑制核酶的活性。

模块合理设计的途径已用于构建由 ATP 控制的几个自身分裂的核酶。依赖 ATP 的别构核酶是由锤头核酶结构域与结合 ATP 的 RNA 结构域构成（图 4.20），许多其他的别构核酶也已合成。某些结合配体的结构域结构诱导契合（往往称适应结合）产生催化结构域相应的结构改变。通过合理设计所产生的大多数别构锤头核酶仅仅是动态范围的一小部分。应用模块合理设计策略工程别构核酶很可能不是产生分子转换的最佳路线。

体外选择策略成功地分离出许多新的核酶和脱氧核酶。其中一个重要的新的别构核酶是由 RNA 随机顺序群体分离出来的。当效应因子加入时，其催化活性调节大约 10 000 倍。RNA 连接酶核酶“L1”是在起始随机顺序库选择产生的[图 4.21(a)]。由于核酶的别构活化产生连接 RNA 的产物，RT-PCR 可以用于扩增核酶，因此反应中的寡核苷酸可以定量[图 4.21(b)]。这样可以把依赖 RNA 的核酶或脱氧核酶组装成基因芯片，直接读出生物样品中的 RNA。

别构核酶产生的主要策略是体外选择和模块合理设计的组合应用。它把催化和受体两个模块由短的随机顺序桥连起来，然后体外选择，由可能的桥顺序鉴定出少数别构核酶。这种策略已产生一系列核酶，例如结合黄素腺嘌呤单核苷酸（FMN）和其他的效应因子的核酶。在某些情况下，这些桥结构域或通讯模块在结合受体结构域之后可以保持它的别构功能。因此，这些小的桥结构域往往起独立模块的作用，它可以产生新的别构核酶。当模块构建没有现有的受体或适体（aptamer）结构域时，别构选择（图 4.22）也用于分离效应因子调节的核酶。环状单核苷酸，抗生素和蛋白质效应因子结合位点是由相对大的随机顺序结构域附加到核酶上，由于大量的可能顺序组合附加到母体核酶结构域上，这样将有许多不同类型的别构结合位点出现在起始群体中。目前别构选择是产生和分离一些核酶最好的策略。

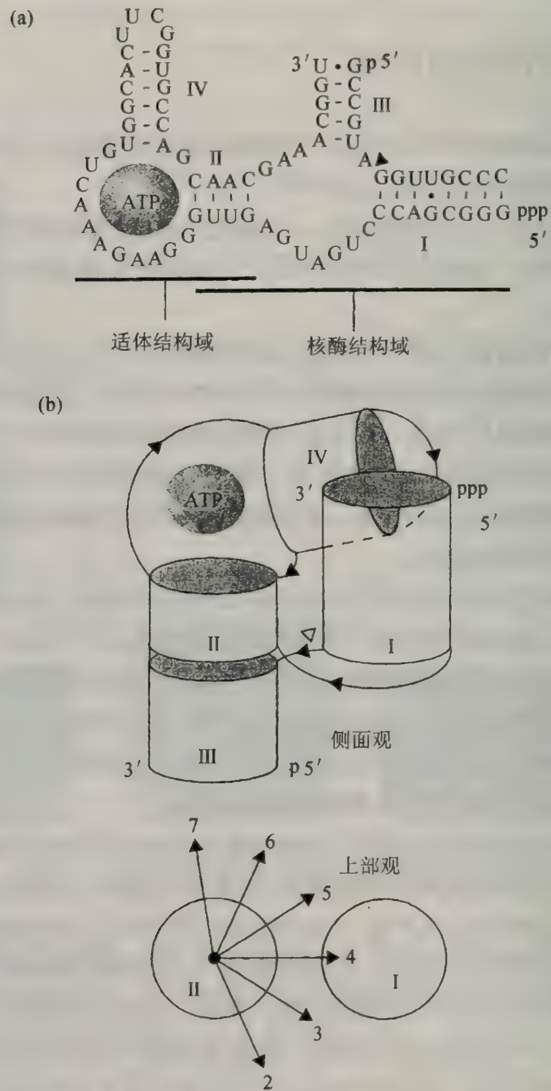


图 4.20 模块合理设计产生依赖 ATP 的别构核酶

(a) 构成物 IV 上方的顺序和二级结构是由锤头核酶结构域和结合 ATP 的 RNA 适体结构域组成。两个结构域共享茎 II 结构单元。(b)依赖 ATP 的别构核酶的三维结构图解。侧面观:椭圆形阴影表示 Watson-Crick 碱基配对单元,箭头代表单股区的极性(5′到 3′)。空箭头指示锤头核酶分裂的位点。上部观:箭头表示茎 IV 同茎 II 中的 2～7 碱基对构成物的方位

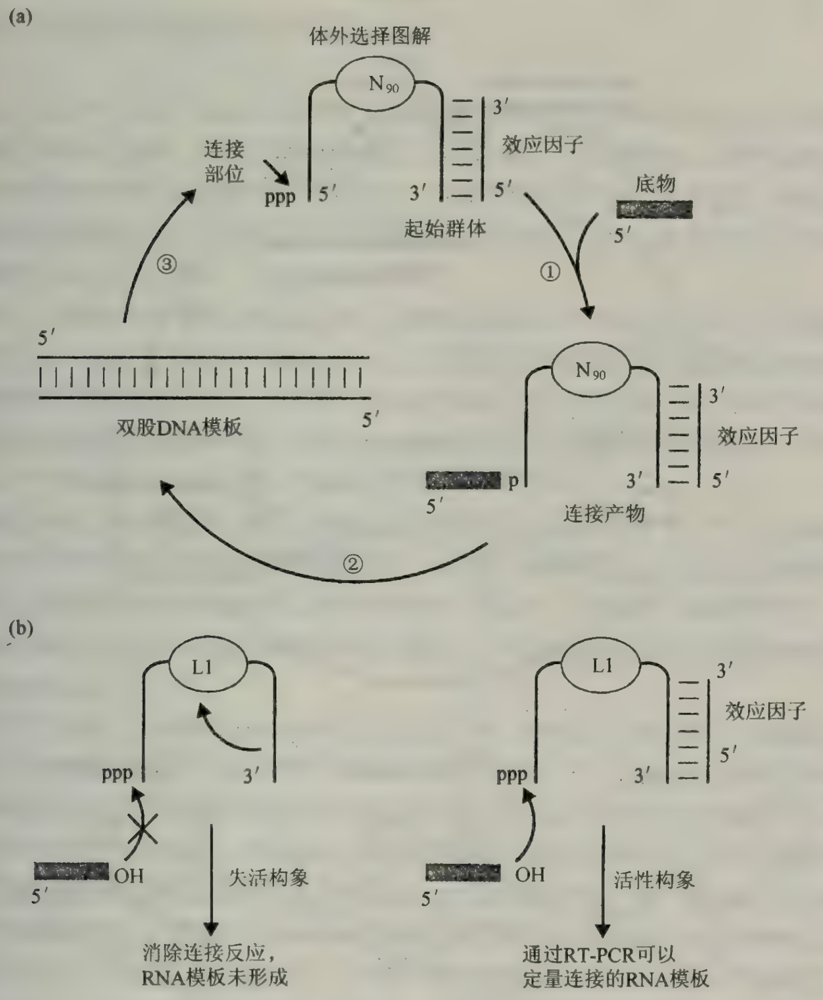


图 4.21　依赖寡核苷酸 RNA 的连接酶核酶的分离和应用

(a)体外选择以 RNA 构成物起始，包括 90 核苷酸随机顺序结构域(N_{90})两侧有 5′和 3′引物结合位点。RNA 转录群体(每个都带有 5′-端三磷酸部位)。①同 DNA 效应因子组合并且在反应条件下同 RNA 底物寡核苷酸温育，底物同 RNA 的 5′端连接；②分离无活性的 RNA 并用 RT-PCR 扩增，形成双股 DNA 模板；③体外转录产生下一个 RNA 群体。L1RNA 连接酶核酶在五轮选择扩增后由群体分离。(b) L1 核酶的别构功能。左：在缺乏效应因子寡核苷酸的情况下，L1 核酶取无活性结构，排除 RNA 连接。如果 RT-PCR 引物只扩增连接的 RNA 产物，那么可能没有扩增信号出现。右：在效应因子寡核苷酸的存在下，L1 核酶的活性大约提高 10 000 倍。连接 RNA 产物水平可以应用 RT-PCR 进行扩增和定量[81]

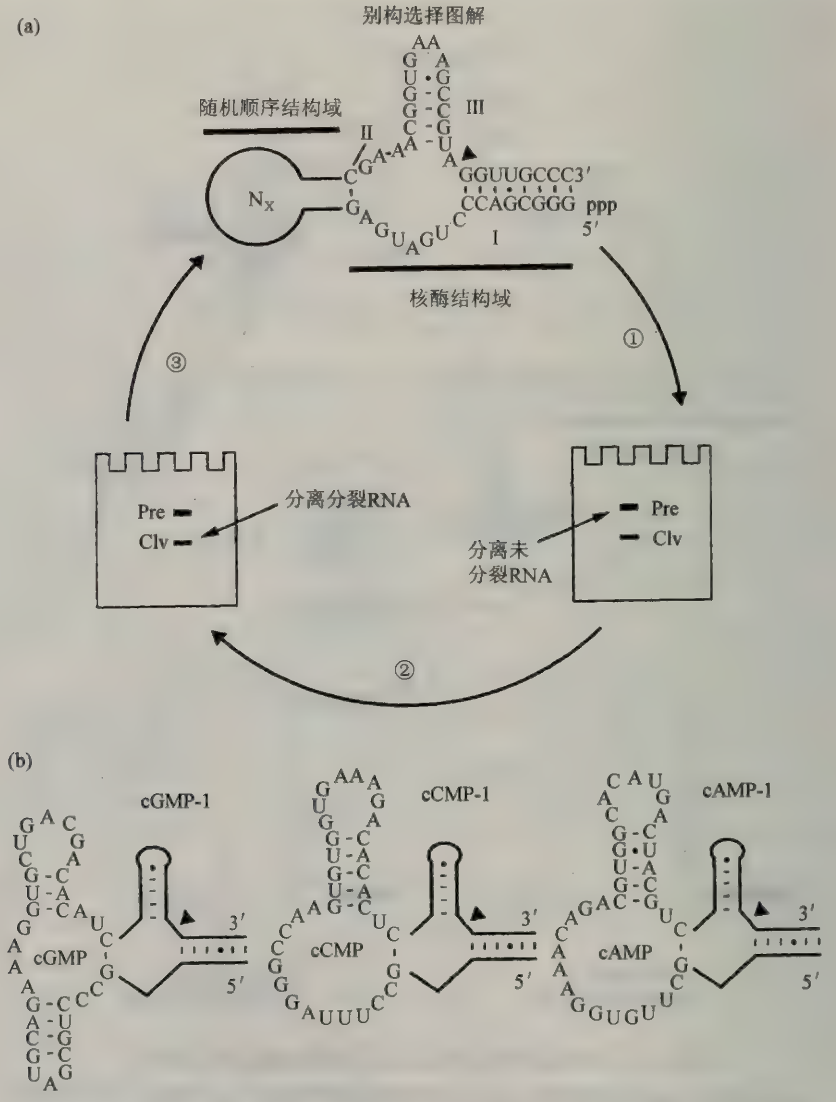


图 4.22 RNA 转换的别构选择

(a) 锤头核酶别构选择图解。选择构成物是由锤头核酶结构域和随机顺序结构域(N_x)组成。①在效应因子缺乏的情况下进行负选择,形成的产物用 PAGE 分离，剩下的未分裂的核酶前体(Pre)由凝胶回收；②在效应因子存在下进行正选择，产物用 PAGE 分离，在这个阶段，被分裂的 RNA(Clv)由凝胶回收。③用 RT-PCR 扩增，随后的 RNA 群体用体外转录制备，并用 PAGE 纯化。(b)由别构选择分离出三个别构锤头核酶。当足够量的3′,5′-环状单核苷酸存在下温育时,每个构成物经效应因子诱导锤头分裂(箭头)

4.4.3　核酶的连续体外进化

在分子水平上研究达尔文的进化论为实验室迅速进化各种生物大分子奠定了思想、概念和理论基础，促进了进化生物技术的发展。由核酸和蛋白质库已经获得结合特殊配体或催化特殊化学转化的分子，但操作通常是分步方式进行，每轮进化都需要纯化选择的分子。如果体外进化能够以连续方式进行，没有任何分离或纯化步骤则为定向进化提供一个理想的模式[82]。

第一个连续体外进化实验是应用 Qβ 噬菌体基因组 RNA，其作为 Qβ 复制酶的底物而涉及一系列转移，其中 RNA 在 NTP 存在下用复制酶扩增，复制混合物小部分转移到新反应器允许进一步扩增，在几次转移之后，进化的 RNA 分子与起始 RNA 比较，更易被复制酶扩增。等温(isothermal)RNA 扩增技术(反向和正向转录偶联)已用于噬菌体启动子单元连续体外进化。

催化 RNA 连接反应的 RNA 分子连续体外进化系统已经建立。类 1 连接酶核酶已作为发展该系统的起点，该酶催化 3′, 5′磷酸二酯键的形成。类 1 连接酶最初由 10^{15}以上随机顺序库中分离出，并且具有惟一形成 3′, 5′磷酸二酯键的型主。类 1 连接酶所催化的反应与聚合酶(蛋白质)类似。类 1 连接酶催化多重模板指导的单核苷酸的加成。

为了应用类 1 连接酶连续体外进化而采用一种嵌合 DNA-RNA 底物，它含 T7RNA 聚合酶(T7RNAP)启动子单元的一股顺序(图 4.23)。除了核酶和底物外，连续体外进化反应混合物中含有逆转录酶，用于启动 cDNA 合成的引物、4 种 dNTP、T7RNAP 和 4 种 NTP。混合物中所有核酶分子都被逆转录。此外，逆转录酶的存在要求核酶先进行连接反应后被逆转录。在正向转录之后，形成的子代 RNA 立即适时的起动另一轮催化 RNA 的连接和选择扩增。这些事件允许在单一反应器中重复发生。然后反应混合物的一小部分转移到新的反应器中进行连续进化。

由上可见，连续体外进化系统，通过单一核酶可以进行多重连续反应，为研究进化的分子群体行为提供了更复杂的模型。类 1 连接酶能够催化模板指导的单核苷酸加成作为构建该系统的基础。

达尔文进化的复杂度使它难于研究进化群体的生化性质的改变，这指出需要简化模型系统，为此发展了生物大分子进化的实验室系统。这些系统或者涉及功能分子分步进化，或者是功能表达和选择扩增一起发生在共同反应器中的连续进化。连续进化已用于进化连接酶核酶，其催化 3 个连续的核苷酸加成反应。为了选择扩增，要求 RNA 分子进行两个单核苷酸加成反应，继之 RNA 连接。这个进化的核酶像 RNA 聚合酶。它支持 RNA 复制酶在 RNA 世界能够进行信息转移的思想。不像以前描述的类聚合酶核酶，其是 5′→3′方向操作。进化的核酶加成单核

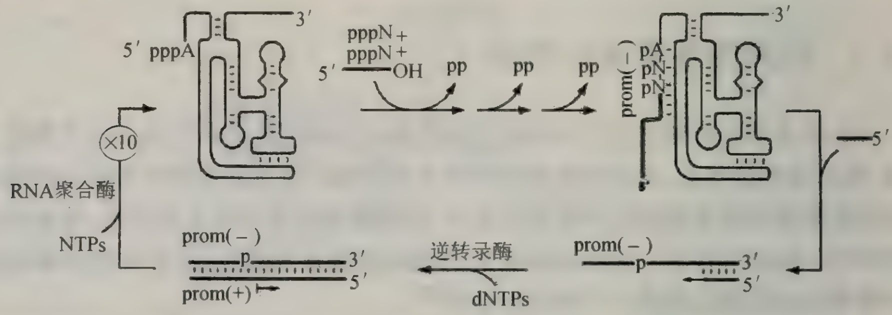


图 4.23　连续体外进化

连接酶的核酶库定向结合嵌合 DNA-RNA 底物(DNA,虚线;RNA,实线),其含有大部分或全部 T7RNAP 启动子单元[prom(－)]。核酶需要通过加入核苷酸完成 T7RNAP 启动子,然后全长启动子与它们特有的 5′端连接。通过逆转录酶将所有 RNA 分子转为 cDNA,给出完全双股的启动子单元,继之通过 T7RNAP 转录产生子代 RNA

苷酸或者 5′→3′或 3′→5′方向。那么是否在 RNA 世界中 RNA 催化 RNA 复制是 5′→3′还是 3′→5′或两者兼之,这是一个有待进一步研究的问题。

4.4.4　核蛋白酶的体外选择

效应因子调节的核酶的活化参数与别构蛋白质相类似。例如 Breaker 等[80]工程一种别构锤头核酶,在 ATP 存在下被抑制 180 倍。Robertson 等[15]工程一个效应因子活化的核酶(连接酶),在茶碱存在下被活化 1600 倍。由随机库选择出的别构域被附加到锤头核酶上,这些结构域通过环状核苷单磷酸使核酶活化 5000 倍。这些结果指出用蛋白质效应因子活化核酶也是可能的。天然核酶的活性依赖蛋白质。例如真细菌催化 tRNA 分裂的 RNaseP 核酶是具有蛋白质(P 蛋白)的复合物,实际上提高其活性。同样,组 I 内含子 ND1 完全依赖于 Cyt18(*Neurospora crassa* 线粒体的酪氨酰 tRNA 合成酶),而内含子 b15 的三级结构是由其同源蛋白质(CBP2)稳定。蛋白质往往作为分子伴侣协助 RNA 分子的折叠。

Robertson 等[83]开发出体外选择核蛋白酶的方法,并鉴定出新的一类核酶,其专一地通过蛋白质因子(例如 Cy518 和溶菌酶)活化几千倍,并且能够特异地识别天然蛋白质结构。该选择方法具有普遍性,可以高通量鉴定并识别蛋白质组中的大部分蛋白质。

为了鉴定依赖肽和蛋白质的核酶,Roberson 等试图用小分子效应因子的策略来体外选择依赖蛋白质的核酶。然而,当结合肽和蛋白质的位点被加到 L1 连接酶茎 C 上时[图 4.24(a)],在同源肽和蛋白质效应因子存在下,很小或没有调节活

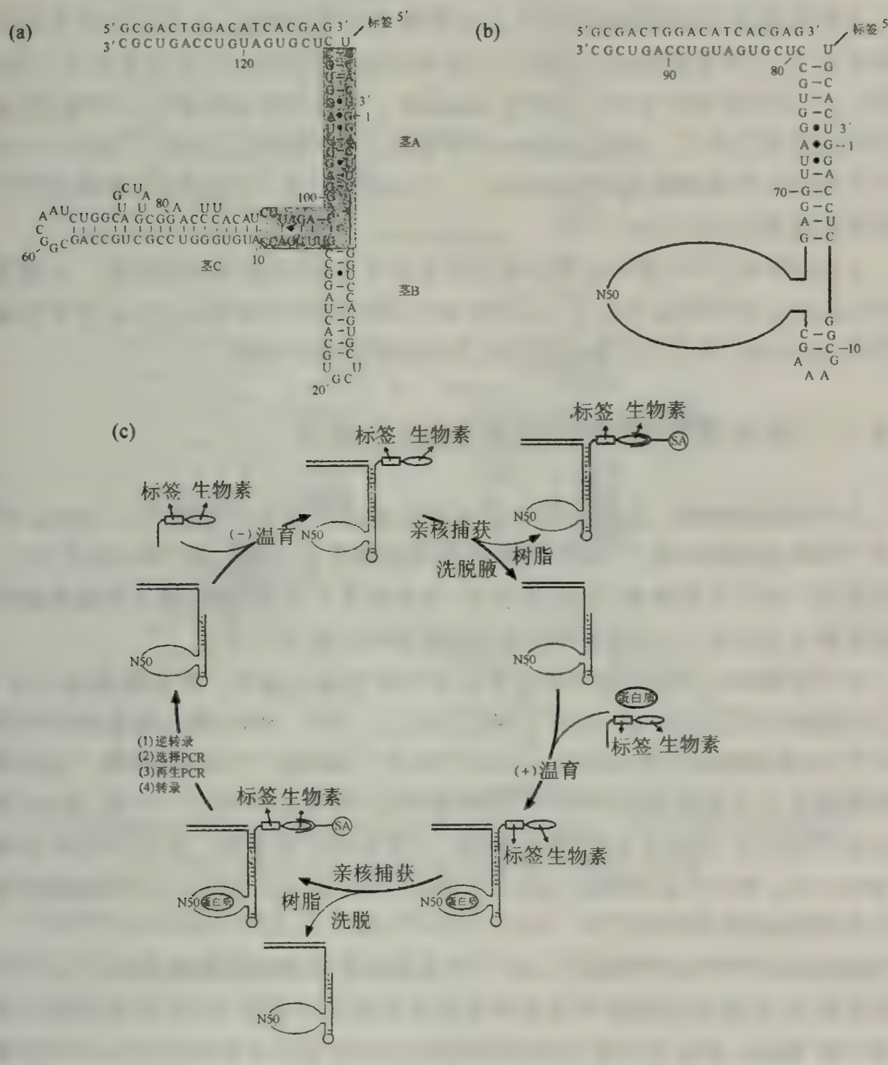


图 4.24　L1 连接酶,L1-N50 库和选择图解

(a)L1 连接酶[15]是库设计的起点，图中指出茎 A、B 和 C，阴影区表示催化核心和连接的接合点。引物结合位点以小写字母表示，斜体字表示活性所需要的寡核苷酸效应因子。黑体表示连接的底物。连接底物的“标签”(tag)是变动的，这里是生物素-$(dA)_{22}$。(b)L1-N50 库含有 50 个随机顺序位置并与核酶核心部分的重叠。茎 B 变小并用稳定的 GNRA 四突环终止，茎 A 的 U5 突变成 C(黑体)同 G69 形成碱基对以增加茎的稳定性。(c)选择图解。RNA 库同生物素化的底物一起温育，并且活性突变体由群体去除。余者再一次在靶蛋白质(Cyt18 或溶菌酶)存在下同生物素化的底物温育。活性突变体由群体中去除，并优先用逆录 PCR 扩增，体外转录[84]

性。同样,当将随机顺序突环引入茎 C 端时,可选择出活化的突变体。

工程依赖蛋白质的核酶不同于工程依赖小分子的核酶,小分子结合在有限的别构位点上,但明显影响核心核酶的二级和三级结构再组织;大效应因子分子(例如蛋白质)很可能与更大的位点结合,在立体上可能抑制催化核心。因此,在选择上必须包括催化核心。为此,Robertson 等根据 L1 连接酶(L1-N50)[15]设计一个库,其中关键催化残基随机化[图 4.24(b)]。L1-N50 库(从 10^{15}起始)对连接活性经历正和负重复选择[图 4.24(c)]。

在天然核蛋白中,蛋白质成分通过稳定活性 RNA 构象体活化核酸。依赖蛋白质的 aptazyme 可以固定在芯片上,特异地识别它们的目标蛋白,可以开发出高通量选择 aptazyme,其允许定量检验蛋白质组和代谢组中的靶。

4.4.5　延伸蛋白质编码的核酶体外选择

非天然氨基酸移入蛋白质可以形成具有新化学性质的材料[85]。Bessho 等[86]描述了特殊选择的核酶,可以把氨基酸转移到特异的 tRNA 的 3′端(图 4.25)。该方法提供一种产生氨酰基 tRNA 的工具,其载有非天然氨基酸,通过无细胞翻译可以最终移入蛋白质中,产生各种用途的新性质的工程蛋白质。

为了控制蛋白质的生物合成需要特异的改变遗传编码。在氨酰基化反应中,每个氨基酸结合在同源 tRNA 的 3′端(图 4.25)。每个 tRNA 都有氨基酸特异的反密码子,当氨基酸结合在特异的 tRNA 时,建立了编码的三联体-氨基酸。这些通用的氨酰基化反应是由氨酰 tRNA 合成酶催化的,每氨基酸对应一个酶,tRNA 合成酶是古代蛋白质,较早出现在“RNA 世界”过渡到蛋白质中的。由于 mRNA 的密码子-tRNA 的反密码子在核糖体上相互作用的结果,结合在 tRNA 上的氨基酸被插入到生长多肽链的一定位置。

氨酰基化特异性是明显的,一旦一个氨基酸被稳定地结合在 tRNA 上,它的命运就已确定,这是由于密码子-反密码子相互作用对结合在 tRNA 的氨基酸性质不敏感。在 Bessho 等工作之前,为改变氨酰基化而发展了 4 个系统(表 4.1),依靠氨酰基化方法,来自这些特异的氨酰基化系统的蛋白质可能是纯的化学实体或密切相关多肽统计上的集合。上述系统中的 3 个可以用于体内,这样,细胞实际上产生含有非天然氨基酸的蛋白质,即细胞面临吸取一定的非天然氨基酸,经一个 tRNA 合成酶作用,破坏特异的 tRNA 合成酶的编辑能力,由于氨酰基化的错误,天然机制被去除,并且产生的突变体 tRNA 合成酶对特殊的非天然氨基酸-tRNA 是特异的。虽然上述 3 个途径已用于体内,它们也可能适于体外制造新的蛋白质。第四个途径只能用于体外,是用化学和酶法把新的氨基酸连接到特异的 tRNA 上,典型的例子是 tRNA 被指定阅读终止密码子(例如 UAG)。

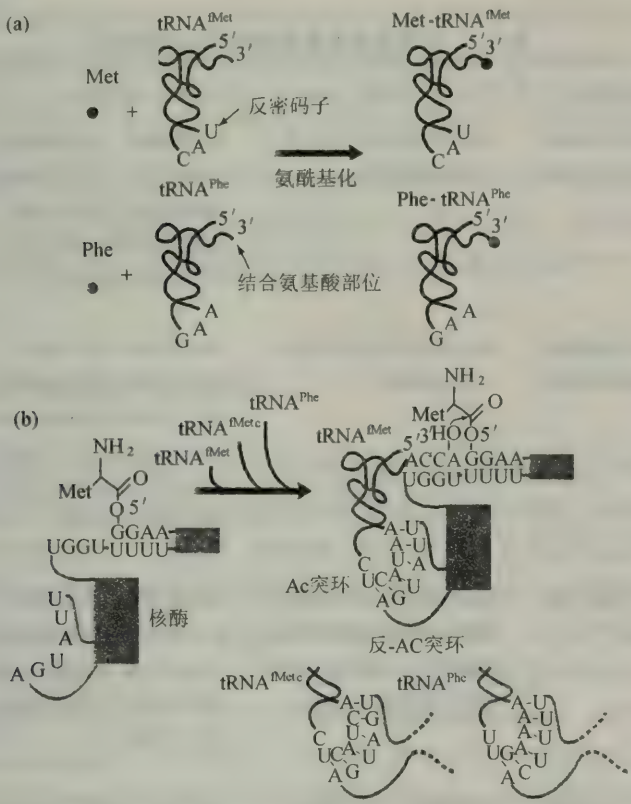


图 4.25　带有反密码子的核酶氨酰基化 tRNA

(a)氨酰基化的图解表示，其中每个氨基酸(●)与 tRNA3′端结合；(b)应用特异的核酶氨酰基化不同的 tRNA。各种核酶仅在反密码子突环上有区别

许多学者认为最早氨酰基化是由核酶催化的。每个核酶都有一个反-反密码突环。这样，载有氨基酸的核酶通过反密码子-反-反密码子互补碱基对与其 tRNA 杂化。很明显，反密码子-反-反密码子相互作用不是核酶的设计部分，而是选择的结果。核酶(按设计)也与所有 tRNA 3′端的单股部分杂化。这样，tRNA 底物被核酶占据两个地方。氨酰基由核酶的 5′-OH 转移到 tRNA 的 3′-OH 给出所要求的氨酰基化 tRNA。

三个不同的 tRNA 已表明特异地同有互补反-反密码子突环的核酶相互作用，并且不与它们的反密码子-反-反密码子配对中错配的其他两个核酶相互作用。有趣的是，在天然发生的系统中，许多 tRNA 合成酶与它们同源 tRNA 结合。在其他的相互作用中，同反密码子三联体和核苷酸特殊的相互作用接近 tRNA 的 3′端。这样，模拟天然合成酶-tRNA 相互作用，Bessho 等选择产生的核酶，重演 RNA-RNA (核酶-tRNA 相互作用)。

表 4.1 产生具有新氨基酸的蛋白质的方法

氨酰基化方法	应 用	评 注	文 献
作用于一定的非天然氨基酸的野生型或突变体 tRNA 合成酶	具有新化学性质的多肽，作为生物材料	体内或体外，产生"统计上的蛋白质"。不能用编辑功能清除非天然氨基酸；非天然氨基酸大规模移人不需要特殊化学方法	[93]
通过破坏 tRNA 合成酶的编辑结构域模糊遗传编码	具有新化学性质的多肽，作为生物材料	体内或体外，产生"统计上的蛋白质"。越过编辑结构域，这样非天然氨基酸的大库可以移人。大规模移人非天然氨基酸，不需要特殊的化学方法	[94]
突变体 tRNA 合成酶合成新氨基酸-tRNA	新氨基酸单一或多重插入到特异多肽的特异位置。在确定的内部位置产生具有定位探针的蛋白质	体内或体外同源蛋白质产物。需要靶蛋白质的基因或 mRNA 有新的第 21 氨基酸的密码子(新氨基酸插入的位置)	[95]
特异 tRNA 的化学氨酰基化	新氨基酸单一或多重插入到特异多肽的特异位置。在确定的内部位置产生具有定位探针的蛋白质	只在体外，基因(或 mRNA)必须在新氨基酸引人的位置改变密码子，允许新的氨基酸插入到靶蛋白的特异位置	[96]
tRNA 特异的核酶	新氨基酸单一或多重插入到特异多肽的特异位置。在确定的内部位置产生具有定位探针的蛋白质	只在体外，核酶反-反密码子的简单改变允许由一个 tRNA 移到另一个 tRNA。允许新氨基酸插入到所有位置，也可以允许单一插入到特殊密码子的位置	[97]

Bessho 等工作的缺点是氨酰基部位由核酶的 5′-OH 转移到 tRNA 的 3′-OH 是固定无效的。由于在 RNA 的 3′端的顺式-二醇(2′-和 3′-OH)较之分离的 5′-OH 产生不太稳定的化学环境。这样，氨酰基 tRNA 的产生需要高浓度的氨酰基化核酶。同样，该方法需要构建荷载的核酶，部分通过化学方法。Bessho 等工作拓展了荷载

具有新氨基酸 tRNA 的可能性,选择出对每个氨基酸和 tRNA 特异的核酶也是可能的。用基本的核酶只需要改变反-反密码子顺序,就可以对特殊的 tRNA 有特异性,不需要其他改变。改变程序性核酶的 tRNA 的专一性似乎比改变氨酰基 tRNA 合成酶的专一性在技术上更容易。

总之,通过体外定向进化可以产生新DNA和RNA分子;可以提供古代RNA进化的模型;可以制造传感器和催化剂用于工业和治疗。拓展核酸顺序空间和扩大核酸功能新方法的开发将增加核酸的结构和功能多样性,增加其应用范围。

4.4.6　体外选择环状核酶

最近的研究工作表明很小的环状单股 DNA(ssDNA)可以被噬菌体和细菌 RNA 聚合酶有效的转录,转录是按滚动环机制进行(图 4.26)。ssDNA 环的大小在 18~110nt,或直径在 2 ~ 10nm。这样的 DNA 纳米环可用于编码连环体(concatamer)或自身分裂产生单体锤头核酶、发夹和肝炎 δ 型主的核酶,但不能复制。纳米环转录起始点在环中任何胞嘧啶处并以 GTP 起始。在某种情况下纳米环的二级结构抑制或提高转录。Kool 等[87]通过体外选择策略选择出约 40nt 顺序作为 DNA 纳米环载体中的启动子,称为伪启动子(pseudopromoter),又称类启动子结构域(promoter-like domain)。它明显小于大肠杆菌中发现的单股启动子。这个启动子可以用于起动 DNA 纳米环载体在大肠杆菌中转录出活性锤头核酶。DNA 纳米环载体库是由一个含 40nt 随机结构域和编码锤头核酶及其分裂底物非随机结构域(63nt)组成的。该库大约有 10^{13}不同的单股环状顺序(图 4.27)。体外选择第 50 轮得到的最佳顺序。

滚动环转录同自身加工结合是核酶合成的有效策略。纳米环载体代表了生物活性 RNA 传递到细菌中一种潜在的有效途径。编码重复(非自身加工)核酶的环状载体可以产生连环体核酶,其靶顺序若是 HIV-1 基因组,这将对艾滋病的防治具有重要作用。

根据酶活性部位柔性的基本理论[88],考虑 RNA 柔性比 DNA 大,我们设计了一个环状 RNA-DNA 酶,它是由 T7 启动子、32ntDNA 和 13ntRNA 酶组成的[89]。

在上述工作基础上,我们考虑核酶缺乏多种催化基团,各种二级结构相互作用(特别是疏水核心)和调节结构域,设计一个环状 RNA-DNA 酶[90]。它是由一个来自锤头核酶的 22nt 催化结构域和一个 55nt DNA 片段组成,后者含两个结合结构域(分别是 9nt 和 6nt)和一个调节结构域(带有 20nt 随机顺序的 40nt),见图 4.28。通过体外选择获得一个环状 RNA-DNA 酶(cRDE),较之锤头核酶,它有较高的活性和稳定性以及对二价金属离子的亲和力。通过引入调节结构域经表观交替变位调

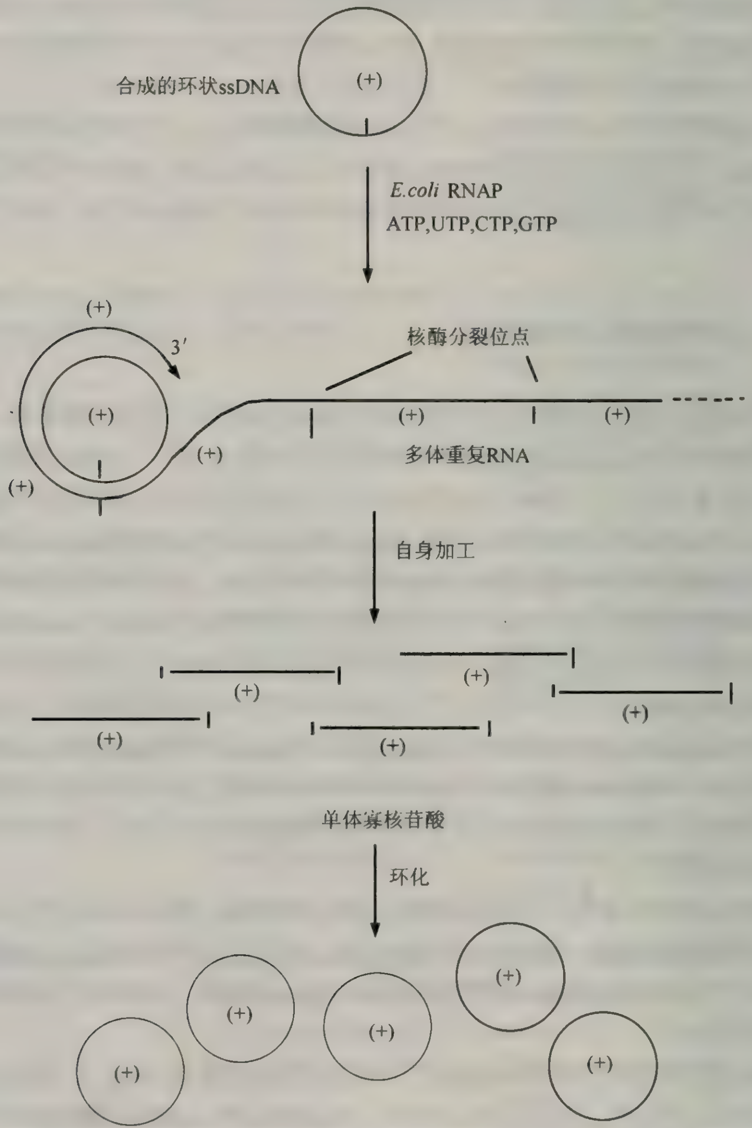


图 4.26　编码发夹核酶的环状 DNA 滚动环转录并自身加工及连接过程

节刚性(对稳定性而言)和柔性(对活性而言)之间的最佳平衡,提高环状 RNA-DNA 酶对环境的适应性。

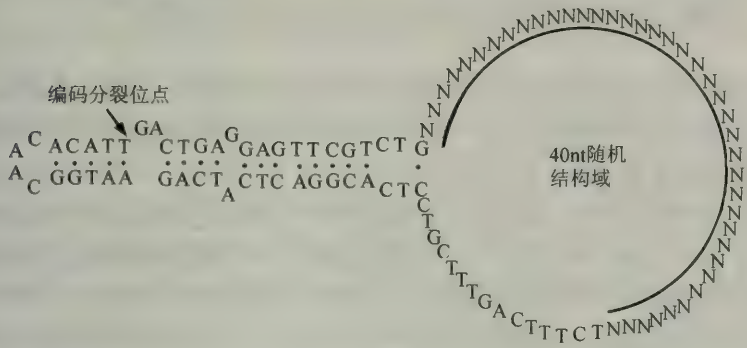


图 4.27　103nt 单股 DNA 纳米环库，其含有 40nt 随机顺序和 63nt 编码锤头核酶及其分裂底物的固定顺序

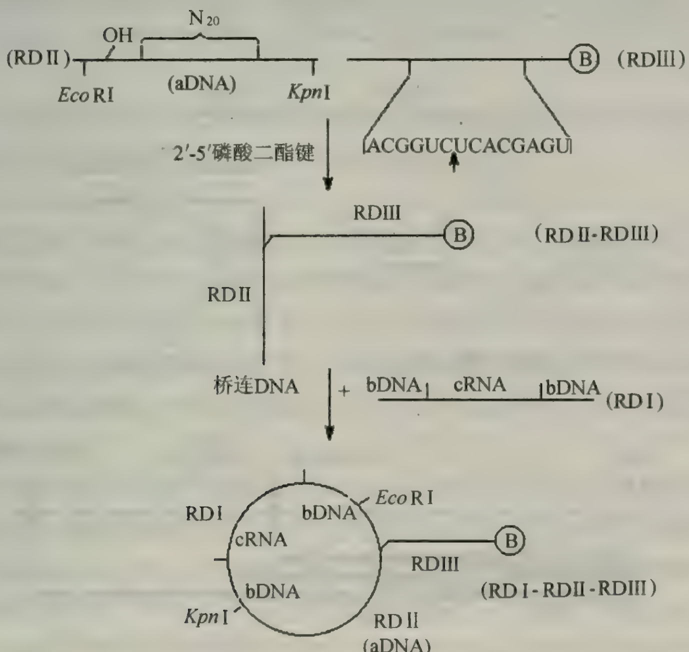


图 4.28　环状 RNA-DNA 酶的设计和构建

—OH 代表 RDII 9 位腺苷的 2′-羟基，用于同 RDIII 的 5′-羟基形成 2′5′-磷酸二酯键；N_{20}表示 20nt 随机顺序；箭头指出环状 RNA-DNA 酶分裂位点；B 表示亲合选择的生物素；aDNA 代表辅助 DNA；bDNA 代表结合 DNA；cRNA 代表催化 RNA

4.4.7　核酶/脱氧核酶的催化潜能和进化策略[91]

虽然 RNA/DNA 已经显示出可以催化许多反应的能力，但毫无疑问蛋白质酶类仍是生物催化剂中的绝对主力。目前已经获得的核酶与蛋白质在催化反应类型和活力上都相差甚远(图 4.29)，这主要有两个方面的原因。一方面，从蛋白质和核酸本身的化学组成来看，蛋白质比核酸具有更大的催化潜能，这表现在蛋白质分子中有许多 RNA/DNA 中所没有的共价催化基团(如半胱氨酸的巯基、丝氨酸的羟基、组氨酸的咪唑基)以及酸碱催化基团(如氨基、羧基、巯基、酚羟基及咪唑等)；蛋白质分子中还可以创造出非极性的活性中心空穴，在这里催化基团被低介电环境所包围，这样底物分子和酶的催化基团之间氢键及静电相互作用被增强，有助于加速酶促反应；蛋白质的骨架的刚性比核酸强，可以折叠成更加紧密稳定的三级结构，有助于形成复杂而稳定催化构象；蛋白质的组成元件有 20 种，同样数量的元件，蛋白质比核酸的组合多样性多出许多数量级，这样从中得到可以催化某一反应

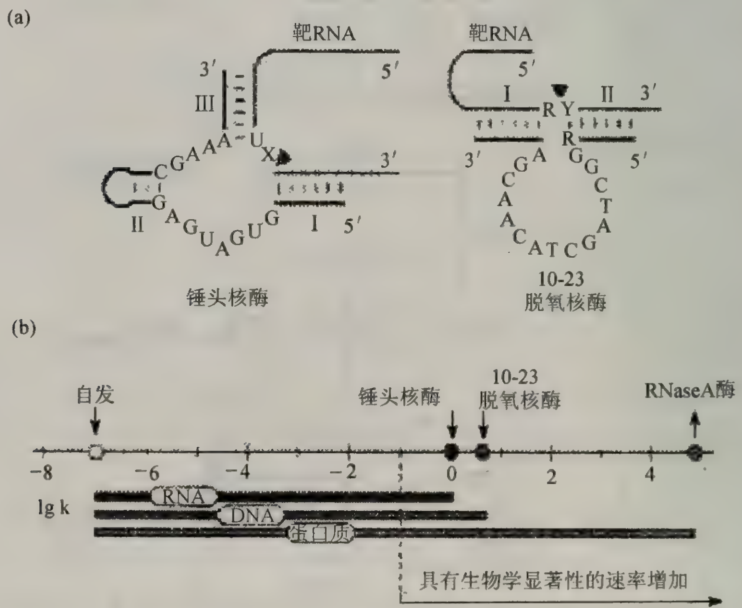


图 4.29　具有切割 RNA 的核酶，脱氧核酶及蛋白质酶(proteinous enzyme)的比较

(a) 是天然锤头核酶和 10-23 脱氧核酶。箭头表示酶催化分裂 RNA 的位点；(b) 是 3 个酶按同一转酯机制催化活性的比较。横坐标表示每个酶催化速率常数(min^{-1})的对数值。带状表示 3 个酶催化速率/RNA 自发分裂速率常数

的顺序的可能性大大增加，同时获得的有催化能力的结构也更为合理。另一方面，蛋白质分子在整个自然界中存在，并经过了千百万年的进化，已经获得十分复杂的结构，而核酶分子仅仅存在于实验室的试管中，才经过短短几年、几个月的分子进化。可以想象，目前 DNA/RNA 酶与蛋白质酶如此大的差距并不只因为它们本身的催化潜能，还有很大一部分原因是前者的许多结构和功能还远没有被开发出来。在自然界中进化的组 I 内含子和 RNase P 就具有了复杂结构，可以加快反应速度达 10^{11} 倍。组 I 内含子在很多方面可以和那些最精密的蛋白质酶相比，它催化完成外显子之间的重组，通过构象的改变连续完成两个不同的反应，并使产物快速释放防止逆反应发生。在实验中进行的定向进化还不能获得如此复杂的(脱氧)核酶，因为目前的体外选择无论在空间上还是在时间上毕竟有限。一个包含 220 核苷酸的核酸分子，共有 10^{132} 种可能的顺序。如果构建一个库，即使每种分子只有一个的话，那么这个库的质量将大约为 10^{110}kg。实际上用于筛选(脱氧)核酶大约只有 10^{15} 种分子，与整个的可能性来比，简直微不足道。通过随机筛选获得的(脱氧)核酶基本上是以比较高的频率出现的，这样才会被 10^{15} 大小的库所包括。催化某一类反应需要具有高度复杂结构的出现频率很低的核酶，通过在实验室中获得的机会比较小。

为了获得催化种类更加多样、性质更加优良的(脱氧)核酶，主要有以下几种思路来克服由于核酸本身以及筛选条件造成的障碍。

(1) 引入其他催化基团

大多数(脱氧)核酶采用金属离子等辅助因子作为催化基团，为了增加(脱氧)核酶的催化潜能，人们设想将额外的功能团引入 RNA/DNA 中以扩增它们结构和功能多样性，例如用氨基酸体外选择以 L-His 为辅助因子催化 RNA 切割反应的脱氧核酶)或其他有机物作为辅助因子。蛋白质酶利用辅酶提供更多共价催化基团和作为活化基团的载体，同样，核酶也可以用这些辅助基团，或者更适合核酶分子的一些辅酶。

还可以直接将一些活性基团引入到 DNA/RNA 分子上，其中修饰核酸开发的方法很有前途。例如对 dATP 和 dUTP 进行化学修饰，分别合成 dATP 的氨基衍生物和 dUTP 咪唑基衍生物，这些衍生物可以被 Taq 酶用作底物完成 PCR 反应。如果将这些衍生物通过 PCR 代替 dATP 和 dUTP 而引入到随机 DNA 分子库中，就可以因为这些额外功能基团的参与而增加了体外筛选出不同活性脱氧核酶的可能性，这些脱氧核酶将具有更高的活性以及更为复杂的结构和功能，同时可以抵抗核酸酶的降解。这个方法与一般的为了提高核酸分子的稳定性而进行化学修饰相比，主要的优点是大大提高脱氧核酶的催化潜能，另外甚至更重要的是这些修饰基团的引入可能通过 PCR 来完成，这样可以实现直接对修饰核酸的筛选，修饰和筛选的统一，使筛选直接针对修饰核酸本身进行，最大限度地开发了修饰核酸的催化

潜能。一般的修饰方法都是对已经获得的核酶或脱氧核酶在某些位置有目的的修饰替换,通过人工的化学合成来得到改造的核酶,这个过程成本比较高,并且无法实现大规模的筛选,改造后的核酶是否具有活性需要逐一来检测。

(2) 扩大筛选的容量

结构复杂的核酶需要更多的保守顺序来维持其特殊空间结构,这样的核酶出现的频率极低,一个很自然的想法就是提高核酸突变库的大小来增加筛选到这些复杂核酶的机会,关键是如何增加突变库的大小。现在采用的从一个随机库出发,每轮用易错 PCR 来增加突变的筛选方法在一定程度上扩大了筛选容量。例如第一轮筛选从一个 10^{15}大小的 N220 随机库开始,以后每一轮通过易错 PCR 在每个 DNA/RNA 分子上引入若干突变,把筛选得到的具有催化潜能的分子库(假设为 10^5)增加了 10^{10}倍,使每轮筛选的突变库都维持在 10^{15}左右,即在筛选能力允许的范围内,又增加了筛选范围。那么经过 4 轮后,所得到的核酶就相当于从一个 10^{45} $(45=15+30)$初始随机库中筛选得到的。当然经过几轮筛选后,得到的突变库内部分子之间,以及和前几轮筛选得到的库之间会有很大的重复,所以映射的初始随机库的容量并不是可以无限增加到包含所有可能的序列。所得到的核酶都是从第一轮筛选到的有催化潜能的分子进化而来的,是在这些分子结构的基础上做进一步细致调整,得到这类结构最优化的催化分子。但第一轮没有获得的一些种类的结构就会被遗漏掉。所以首轮的筛选库的容量十分重要,提高待筛选的随机库的浓度和体积,或者增加选择系统的敏感性,可以增加首轮随机库的容量,但所获得的库的增加是有限的,而且实验成本也大大增加。总之,很难或者是不可能)将随机库的容量增加到可以筛选得到任何复杂核酶的大小。

(3) 模拟自然进化的合理性设计

自然界中,蛋白质酶的自然进化也不是单纯在蛋白质分子的随机库中进行的,整个自然界可以容纳的随机库的大小与整个随机库的可能性相比也是十分有限的。核酶的体外选择进化也可以通过模拟自然进化来突破库容量的限制。蛋白质的自然进化主要在选择压力下进行突变,重组和模块组装。对于前两种模式,可以通过易错 PCR 和 DNA 改组等手段进行模拟。模块组装对于相对分子质量大、结构和功能复杂的蛋白质最为重要,是突破了筛选库容量限制的进化方式。通过基因的重复、颠倒、插入、缺失或外显子重组等,在分子内实现了若干已经存在的结构和功能单元之间的组装,可以是来自同一分子内的模块在分子内的重复、重排以及颠倒等,也可是来自不同分子的模块的重新组合。分子间也通过相互作用形成寡聚蛋白,甚至更为复杂的细胞器(例如核糖体)。所以复杂的 RNA 分子可以通过模拟自然界中蛋白质进化的方式对筛选获得的结构和功能模块进行组装来实现。将这些模块以不同方式结合在一起,通过不同模块之间的随机连接,并在连接的部分插入随机顺序,构建出一个模块组装的随机库,然后进行筛选,获得所需要的模块

组合方式。这种进化方式归根结底也是扩大了筛选库的大小，假如一个从 10^{13} 大小的随机库中经过若干轮易错 PCR 筛选得到的模块相当于从大小为 10^{40} 的库，5 个这样的模块经过重组获得的核酶，就相当于从容量为 10^{200} 的随机库中筛选得到的，更多模块则有更多的重组可能，这样定向进化的核酶的复杂程度就会大大增加。

剪切型核酶就可以看成是两种简单模块组合而成，底物结合模块和催化模块，通过更换与催化模块连接的底物结合模块，可以对不同的底物进行切割。对于催化其他反应类型的核酶也可以从底物结合的 aptamer 模块（库）与催化这类反应的模块（库）进行组合筛选，还可以组装上利用各种辅助因子（ATP，氨基酸等）的 aptamer 模块（库），以获得额外的催化基团、反应中间体载体或者能量等。另外也可以像蛋白质具有调控部位一样，设计调控模块，这就是别构核酶。已经设计了一系列有黄素 aptamer 序列的别构锤头型核酶，它包含特异结合黄素的位点并受到黄素的调控。黄素单核苷酸存在的情况下，通过改变核酶的构象，提高了催化效率，其中一种别构核酶在黄素单核苷酸饱和时，K_{cat} 增加了 10 倍（不知道是否也跟 FMN 提供活性基团有关）。还有许多相似的报道，这些别构核酶都是携带与活性部位分开的效应因子结合部位，配体与这个别构部位结合，诱导相邻核酶结构域的构象变化，以提高或抑制催化机制。

核酶的体外选择尚处在幼年时期，今后随着人们获得的核酶种类、掺入核酶中的活性基团以及核酶可以利用的辅助因子的增多，并通过对各种模块的重组筛选，核酶分子将在实验室中不断进化，相信会出现更多种类、结构和功能更为复杂的核酶。

核酶的研究方向中有两个领域值得特别关注：一个是蛋白质同核酶的相互作用；另一个是核酶的结构生物学[92]。

蛋白质能够提高或扩展中心 RNA 的活性。核糖体和剪接小体是最重要的两个系统，也是最重要的核酶领域。通过这些蛋白质的机制和结构相互作用促进 RNA 催化反应。核酶的结构生物学，即在原子水平上解析大的核酶（组 I 或 II 内含子，或 RNaseP）是另一重要领域。核酶的结构生物学将越来越多地集中在体外选择的核酶和天然催化核蛋白颗粒。它包括 RNaseP 全酶，各种组 I 内含子-成熟酶蛋白复合物、剪接小体的成分以及核蛋白酶（例如端粒酶）。这些将是核酶领域今后要做的主要工作。

参　考　文　献

1　Moore P B，Steitz T A. The involvement of RNA in ribosome function. Nature，2002，418：229～235

2　Ban N，Nissen P，Hansen J，*et al*. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. Science，2000，289：905～920

3 Wimberly B T, Brodersen D E, Clemons W M, *et al*. Structure of the 30S ribosomal subunit. Nature, 2000, 407: 327 ~ 339

4 Yusupov M, Yusupova G Z, Baucom A, *et al*. Crystal structure of the ribosome at 5.5 Å resolution. Science, 2001, 292: 883 ~ 896

5 Joyce G F. The antiquity of RNA-based evolution. Nature, 2002, 418: 214 ~ 221

6 Unrau P J, Bartel D P. RNA-catalysed nucleotide synthesis. Nature, 1998, 395: 260 ~ 263

7 Johnston W K, Unrau P J, Lawrence M S, *et al*. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. Science, 2001, 292: 1319 ~ 1325

8 Lee N, Bessho Y, Wei K, *et al*. Ribozyme-catalyzed tRNA aminoacylation. Nature Struct Biol, 2000, 7: 28 ~ 33

9 Zhang B, Cech TR. Peptide bond formation by *in vitro* selected ribozymes. Nature, 1997, 390: 96 ~ 100

10 Bartel D P, Szostak J W. Isolation of new ribozymes from a large pool of random sequences. Science, 1993, 261: 1411 ~ 1418

11 Ekland E H, Szostak J W, Bartel D P. Structurally complex and highly active RNA ligases derived from random RNA sequences. Science, 1995, 269: 364 ~ 370

12 Eklan E H, Bartel D P. RNA-catalysed RNA polymerization using nucleoside triphosphates. Nature, 1996, 382: 373 ~ 376

13 Jaeger L, Wright M C, Joyce G F. A complex ligasc ribozyme evolved *in vitro* from a group I ribozyme domain. Proc Natl acad Sci USA, 1999, 96: 14712 ~ 14717

14 McGinness K E, Joyce G F. RNA-catalyzed RNA ligation on an external RNA template. Chem Biol, 2002, 9: 297 ~ 307

15 Robertson M P, Ellington A D. *In vitro* selection of an allosteric ribozyme that transduces analytes into amplicons. Nat Biotechnol, 1999, 17: 62 ~ 66

16 Rogers J, Joyce G F. The effect of cytidine on the structure and function of an RNA ligase ribozyme. RNA, 2001, 7: 395 ~ 404

17 Lorsch J, Szostak J W. *In vitro* evolution of new ribozymes with polynucleotide kinase activity. Nature, 1994, 371: 31 ~ 36

18 Hager A J, Szostak J W. Isolation of novel ribozymes that ligate AMP-activated RNA substrates. Chem Biol, 1997, 4: 607 ~ 617

19 Lohse P A, Szostak J W. Ribozyme-catalysed amino-acid transfer reactions. Nature, 1996, 381: 442 ~ 444

20 Jenne A, Famulok M. A novel ribozyme with ester transferase activity. Chem Biol, 1998, 5: 23 ~ 34

21 Wilson C, Szostak J W. *In vitro* evolution of a self-alkylating ribozyme. Nature, 1995, 374: 777 ~ 782

22 Wecker M, Smith D, Gold L. *In vitro* selection of a novel catalytic RNA: characterization of a sulfur alkylation reaction and interaction with a small peptide. RNA, 1996, 2: 982 ~ 994

23 Tarasow T M, Tarasow S L, Eaton B E. RNA-catalysed carbon-carbon bond formation. Nature, 1997, 389: 54 ~ 57

24 Seeling B, Jäschke A. A small catalytic RNA motif with Diels-Alderase activity. Chem Biol, 1999, 6: 167 ~ 176

25 Freeland S J, Knight R D, Landweber L F. Do proteins predate DNA. Science, 1999, 286: 690 ~ 692

26 Strubbe J, Ge J, Yee C S. The evolution of ribonucleotide reduction revisited. Trends Biochem Sci, 2001, 26: 93 ~ 99

27 Kumar R K, Yarus M. RNA-catalyzed amino acid activation. Biochemistry, 2001, 40: 6998 ~ 7004

28 Illangasekare M, Sanchez G, Nickles T, *et al*. Aminoacyl-RNA synthesis catalyzed by an RNA. Science, 1995,

267: 643 ~ 647

29 Illangasekare M, Yarus M. Small-molecule-substrate interactions with a self-aminoacylating ribozyme. J Mol Biol, 1997, 268: 631 ~ 639

30 Saito H, Kourouklis D, Suga H. An *in vitro* evolved precursor tRNA with aminoacylation activity. EMBO J, 2001, 20: 1797 ~ 1806

31 Nissen P, Hansen J, Ban N, *et al*. The structural basis of ribosome activity in peptide bond synthesis. Science, 2000, 289: 920 ~ 930

32 Fontana W, Schuster P. Continuity in evolution: on the nature of transitions. Science, 1998, 280: 1451 ~ 1455

33 Schultes E A, Bartel D P. One sequence, two ribozymes: implications for the emergence of new ribozyme folds. Science, 2000, 289: 448 ~ 452

34 Szostak J W, Bartel D P, Luisi P L. Synthesizing life. Nature, 2001, 409: 387 ~ 390

35 Cech T R, Zaug A J, Grabowski J. *In vitro* splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence. Cell, 1981, 27: 487 ~ 496

36 Haseloff J, Gerlach W L. Simple RNA enzyme with new and highly specific endoribonuclease activities. Nature, 1988, 334: 585 ~ 591

37 Hendry P, McCall M. Unexpected anisotropy in substrate cleavage rates by asymmetric hammerhead ribozymes. Nucleic Acids Research, 1996, 24: 2679 ~ 2684

38 Pleg H W, Flaberty K M, Mckay D B. Three-dimensional structure of a hammerhead ribozyme. Nature, 1994, 372: 68 ~ 74

39 Kowabara T, Warashina M, Taira K. Allosteric controllable maxizymes cleave mRNA with high efficiency and specificity. Trends Biotechnol, 2000, 18: 462 ~ 468

40 Lott W B, Pontius B W, von Hippel P H. A two-metal ion mechanism operates in the hammerhead ribozyem-mediated cleavage of an RNA substrate. Proc Natl Acad Sci USA, 1998, 95: 542 ~ 547

41 Warashina M, Takagi Y, Stec W J, *et al*. Differences among Mechanisms of ribozymecatalyzed reactions. Curr Opin Biotechnol, 2000, 11: 354 ~ 362

42 Berzal-Herranz A, Joseph S, Chowrira B M, *et al*. Essential nucleotide sequences and secondary structure elements of the hairpin ribozyme. EMBO J, 1993, 12: 2567 ~ 2573

43 Pace N R, Smith D. Ribonuclease P: function and variation. J Biol Chem, 1990, 265: 3587 ~ 3590

44 Perreault J P, Altman S. Important 2′-hydroxyl groups in model substrates for M1 RNA, the catalytic RNA subunit of RNase P from *Escherichia coli*. J Mol Biol, 1992, 226: 399 ~ 409

45 Cobaleda C, Sanchez-Garcia, I. RnaseP: from biological function to biotechnological applications. Trends in Biotechnology, 2001, 19: 406 ~ 411

46 Cech T R, Bass B L. Biological catalysis by RNA. Annu Rev Biochem, 1986, 55: 599 ~ 629

47 张今,张红缨,李青山等. 核酸结构与动力学导论. 北京:科学出版社,1995, 155 ~ 166

48 张今,曹淑桂,罗贵民,张学忠,李正强等编著. 分子酶学工程导论. 北京:科学出版社,2003

49 Breaker R R, Joyce G F. A DNA enzyme that cleaves RNA. Chem Biol, 1994, 1: 223 ~ 229

50 Sugimoto N, Okumoto Y. Development of a short Ca^{2+}-dependent deoxyribozyme with RNA cleavage activity. Nucleic Acids Symp Ser, 1999, 42: 281 ~ 282

51 Gourlain T, Sidorov A, Mignet N, *et al*. Enhancing the catalytic repertoire of nucleic acids II. Simultaneous incorporation of amino and imidazolyl functionalities by two modified tripHospHates during PCR. Nucleic Acids Research, 2001, 29: 1898 ~ 1905

52 Santoro S W, Joyce G F. A general purpose RNA cleavaging DNAzyme. Proc Natl Acad Sci USA, 1997, 94: 4262 ~ 4266

53 Santoro S W, Joyce G E. Mechanism and utility of an RNA-cleaving DNA enzyme. Biochemistry, 1998, 27: 13330 ~ 13342

54 Li J, Zheng W, Kwon A H, *et al*. *In vitro* selection and characterization of a highly efficient Zn(Ⅱ)-dependent RNA-cleaving deoxyribozyme. Nucleic Acids Research, 2000, 28: 481 ~ 488

55 Carmi N, Shultz L A, Breaker R R. *In vitro* selection of self-cleaving DNAs. Chem Biol, 1996, 3: 1039 ~ 1046

56 Carmi N, Balkhi S R, Breaker R R. Cleaving DNA with DNA. Proc Natl Acad Sci USA, 1998, 95: 2233 ~ 2237

57 Feldman A R, Sen D. A new and efficient DNA enzyme for the sequence-specific cleavage of RNA. J Mol Biol, 2001, 313: 283 ~ 294

58 Hjizntoniou E, Iseki S, Uney J B, *et al*. DNAzyme-mediated cleavage of Twist transcripts and increase in cellular apoptosis. Biochem Biophys Res Commun, 2003, 300: 178 ~ 181

59 Mei S H, Liu Z, Brennan J D, *et al*. An efficient RNA-cleaving DNA enzyme that synchronizes catalysis with fluorescence signaling. J Am Chem Soc, 2003, 125: 412 ~ 420

60 Flynn-Charlebois A, Wang Y, Prior T K, *et al*. Deoxyribozymes with 2′-5′ RNA ligase activity. J Am Chem Soc, 2003, 125: 2444 ~ 2454

61 Bittker J A, Phillips K J, Liu D R. Recent advances in the *in vitro* evolution of nucleic acids. Current Opinion in Chemical Biology, 2002, 6: 367 ~ 374

62 Johnston W K, Unrau P J, Lawrence M S, *et al*. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. Science, 2001, 292: 1319 ~ 1325

63 Unrau P J, Bartel D P. RNA-catalysed nucleotide synthesis. Nature, 1998, 395: 260 ~ 263

64 Li Y, Liu Y, Breaker R R. Capping DNA with DNA. Biochemistry, 2000, 39: 3106 ~ 3114

65 Schoning K, Scholz P, Guntha S, *et al*. Chemical etiology of nucleic acid structure: the alpha-threofuranosyl-(3′→2′)oligonucleotide system. Science, 2000, 290: 1347 ~ 1351

66 Salehi-Ashtiani K, Szostak J W. *In vitro* evolution suggests multiple origins for the hammerhead ribozyme. Nature, 2001, 414: 82 ~ 84

67 Schultes E A, Bartel D P. One sequence, two ribozymes: implications for the emergence of new ribozyme folds. Science, 2000, 289: 448 ~ 452

68 沈倍奋主编. 分子文库. 北京:科学出版社,2001. 116 ~ 132

69 Tarasow T M, Tarasow S L, Eaton B E. RNA-catalysed carbon-carbon bond formation. Nature, 1997, 289: 54 ~ 57

70 Perrin D M, Garestier T, Helene C. Bridging the gap between proteins and nucleic acids: a metal-independent RNase A minic with two protein-like functionalities. J Am Chem Soc, 2001, 123: 1556 ~ 1563

71 Gourlain T, Sidorov A, Mignet N, *et al*. Enhancing the catalytic repertoire of nucleic acids. Ⅱ. Simultaneous incorporation of amino and imidazolyl functionalities by two modified triphosphates during PCR. Nucleic Acids Research, 2001, 29: 1898 ~ 1905

72 Thum O, Jager S, Famulok M. Functionalized DNA: a new replicable biopolymer. Angew Chem Int Ed Engl, 2001, 40: 3990 ~ 3993

73 Vaish N K, Fraley A W, Szostak J W, *et al*. Expanding the structural and functional diversity of RNA: analog uridine triphosphates as candidates for *in vitro* selection of nucleic acids. Nucleic Acids Research, 2000, 28: 3316 ~ 3322

74 Tang Y, Tirrell D A. Biosynthesis of a highly stable coiled-coil protein containing hexafluoroleucine in an engi-

neered bacterial host. J Am Chem Soc, 2001, 123: 11089 ~ 11090

75 Wang L, Brock A, Herberich B, *et al*. Expanding the genetic code of *Escherichia coli*. Science, 2001, 292: 498 ~ 500

76 Liu D R, Magliery T J, Pastrnak M, *et al*. Engineering a tRNA and aminozcyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins *in vivo*. Proc Natl Acad Sci USA, 1997, 94: 10092 ~ 10097

77 Maier TH. Semisynthetic production of unnatural L-alpha-amino acids by metabolic engineering of the cysteine-biosynthetic pathway. Nat Biotechnol, 2003, 21: 422 ~ 427

78 Thorson J S, Cornish V W, Barrett J E, *et al*. Abiosynthetic approach for the incorporation of unnatural amino acids into proteins. Methods Mol Biol, 1998, 77: 43 ~ 73

79 罗贵民主编. 酶工程. 北京:化学工业出版社,2002

80 Breaker R R. Engineered allosterie riboizymes as biosensor components. Current opinion in Biotechnology, 2002, 13: 31 ~ 39

81 Robertson M P, Ellington A D. *In vitro* selection of an allosteric ribozyme that transduces analytes to amplicons. Nat Biotechnol, 1999, 17: 62 ~ 66

82 McGinness K E, Wright M C, Joyce G F. Continuous *in vitro* evolution of a ribozyme that catalyzes three successive nucleotidyl addition rections. Chem Biol, 2002, 9: 585 ~ 596

83 Robertson M P, Ellington A D. *In vitro* selection of nucleoprotein enzymes. Nat Biotechnol, 2001, 19: 650 ~ 655

84 Robertson M P, Ellington A D. *In vitro* selection of an allosteric ribozyme that transduces analytes to amplicons. Nat Biotechnol, 1999, 17: 62 ~ 66

85 Tamura K, Schimmel P. Ribozyme Programming extends the protein code. Nat Biotechnol, 2002, 20: 669 ~ 670

86 Bessho Y, Hodgson D R H, Suga H. A tRNA aminoacylation system for non-natural amino acids based on a programmable ribozyme. Nat Biotechnol, 2002, 20: 723 ~ 728

87 Ohmich T, Maki A, Kool E T. Efficient bacterial transcription of DNA nanocircle vectors with optimized single-stranded promoters. Proc Natl Acad Sci USA, 2002, 99: 54 ~ 59

88 Tsou C L. Conformational flexibility of enzyme active sites. Science, 1993, 262: 380 ~ 381

89 Zhang J, Zhang H Y, Wang X P. The design and construction of a new nucleozyme. Annals New York Academy of Science, 1998, 864: 636 ~ 639

90 Kong X D, Zhu S Z, Zhang J, *et al*. A circular RNA-DNA enzyme obtained by *in vitro* selection. Biochem Biophys Res Commun, 2002, 292: 1111 ~ 1115

91 Breaker R R. Tech Sight. Molecular biology. Making catalytic DNAs. Science, 2000, 290: 2095 ~ 2096

92 Cech T R. Ribozyme mechanisms and folding. Biochemical Society Transaction, 2002, 30: 1162 ~ 1166

93 van Hest J C M, Kiick K L, Tirrell D A. Efficient incorporation of unsaturated methionine analogues into proteins *in vivo*. J Am Chem Soc, 2000, 122: 1282 ~ 1288

94 Doring V H, Mootz L A, Nagle T L, *et al*. Englarging the amino acid set of *Escherichia coli* by infiltration of the valine-coding pathway. Science, 2001, 292: 501 ~ 504

95 Wang L, Brock A, Herberich B, *et al*. Expanding the genetic code of *Escherichia coli*. Science, 2001, 292: 498 ~ 500

96 Cornish V W, Mendel D, Schultz P G. Probing protein structure and function with an expanded genetic code. Angew Chem Int Ed Engl, 1995, 34: 621 ~ 633

97 Bessho Y, Hodgson D R, Suga H. A tRNA aminoacylation system for non-natural amino acids based on a programmable ribozyme. Nat Biotechnol, 2002, 20: 723 ~ 728

第 5 章　蛋白质酶定向分子进化

5.1 引　　言

虽然酶仅仅是由 20 种天然氨基酸组成的，但它涉及生命的各个方面。酶的多样性来自其结构和功能的多样性，而这种多样性由各种氨基酸聚合成不同的组成顺序和长度来实现。在折叠成复杂的三维结构时，这些大分子转为选择性和专一性的催化剂，催化各种化学反应。

对酶分子的研究可以分为认识和改造两个方面，前者是利用各种生物化学、晶体学、光谱学等方法对天然酶或其突变体进行研究，获得酶分子特征、空间结构、结构和功能之间关系以及氨基酸残基功能等方面的信息，并以此为依据对酶分子进行改造，称为酶分子的合理设计；与此相对应，不需要准确的酶分子结构信息而通过随机突变和筛选等方法对其进行改造，则称为酶分子的非合理设计。非合理设计的实用性较强，往往可以通过随机产生的突变，改进酶的特性。

酶定向分子进化（directed molecular evolution）属于蛋白质的非合理设计，它不需事先了解酶的空间结构和催化机制，人为地创造特殊的进化条件，模拟自然进化机制（随机突变、基因重组和自然选择），在体外改造酶基因，并定向选择（或筛选）出所需性质的突变酶。

天然酶通常缺乏商业价值的催化功能及其他性质。因此对天然酶分子水平上的改造显得十分重要。天然酶在自然条件下已经进化了千百万年，但是酶分子仍然蕴藏着巨大的进化潜力，这是酶体外定向进化的基本先决条件。酶分子存在着进化的潜力主要原因如下：①天然酶在生物体内存在的环境与酶的实际应用环境不同。一个比较好的例子是把枯草杆菌蛋白酶 E 应用于非水相（二甲基甲酰胺，DMF），催化肽合成反应，自然生理条件下进化得比较完善的枯草杆菌蛋白酶 E 由于没有接触过非水环境，因此其活力和稳定性不适应在有机相中完成催化反应。这就为该酶在新的筛选条件下（有机相中）提供了适应该条件的进化空间。②实际应用中，总是期待酶的活力和稳定性越高越好，这样可以加快反应速度、提高酶的利用率、降低反应成本，但在生物体内更重要的是各种生物分子之间协同作用，作为一个整体去适应环境。生物对环境适应的进化主要不是表现为某个酶分子的活力和稳定性的不断提高，而是在于整体的适应能力、调控能力的增强。对于某个酶分子来说，其活力可以受到调节部位的调节，含量可以受到基因表达的调控，而当其酶活力和稳定性已经超过了满足整个体系在环境中生存的需求时，它们的提高

就显得没有必要了,即失去了进化的筛选压力,因而进化的机会很有限,这也为体外定向进化留下了很大的进化空间。③某些酶或蛋白质待进化的性质不是其在生物体内所涉及的,例如对蛋白质药物改造消除其不良反应,这部分性质的改善有着很大的进化潜力。

酶定向分子进化是从一个或多个已经存在的亲本酶(天然的或者人为获得的)出发,经过基因的突变和重组,构建一个人工突变酶库,通过筛选最终获得预先期望的具有某些特性的进化酶。我们以对单一酶分子基因进行定向进化为例,来说明酶分子定向进化的基本实验路线。在待进化酶基因的 PCR 扩增反应中,利用 Taq DNA 多聚酶不具有 3′→5′校对功能,并控制突变库的大小使其与特定的筛选容量相适合,选择适当条件以较低的比率向目的基因中随机引入突变,进行正向突变间的随机组合以构建突变库,凭借定向的选择(或筛选)方法,选出所需性质的优化酶,从而排除其他突变体。也就是说,定向进化的基本规则是“获取你所筛选的突变体”。简言之,定向进化 = 随机突变 + 正向重组 + 选择(或筛选)。与自然进化不同,前者是人为引发的,后者相当于环境作用于突变后的分子群体,相当于通过选择某一方向的进化而排除其他方向突变的作用。酶定向分子进化过程完全是在人为控制下进行的,是酶分子朝向人们期望的特定目标进化。

酶的功能多样性,专一性和效率吸引科学家和工程师通过蛋白质工程制造生物催化剂,以满足实践需要。蛋白质结构和功能的设计促进蛋白质顺序空间的定向拓展。天然蛋白质进化机制使蛋白质功能适应不断变化的环境。因此了解新蛋白质结构和功能出现的原理和动力,可以提供蛋白质工程的指南。本章重点讨论天然蛋白质进化的机制,进化酶的结构和功能以及定向进化的应用。

5.2　自然界蛋白质进化机制

5.2.1　基因复制

基因复制是蛋白质进化的基本机制之一[1,2]。当染色体 DNA 复制或者减数分裂时发生不等交换,可以产生基因的复制,或者串联重复或者产生母体顺序两个分离的拷贝(图 5.1)。在上述情况下,各种拷贝面临 3 种可能的结果:①通过突变积累失活;②不受功能限制,局部突变积累导致功能分化和新功能的进化;③两个基因都保持野生型活性。然而,蛋白质结构和功能对突变非常敏感,通过随机漂移复制趋异很可能造成酶的失活。但模型研究指出在两个基因复制中互补简并突变实际上有利于宿主。两个拷贝延长保守时间增加了进化新功能的可能性。

新功能蛋白质进化的另一种观点包括基因募集(recruitment)或基因分配。在这个模式中,单一基因以双功能周期性进行复制,一个拷贝保持祖先功能,而另一

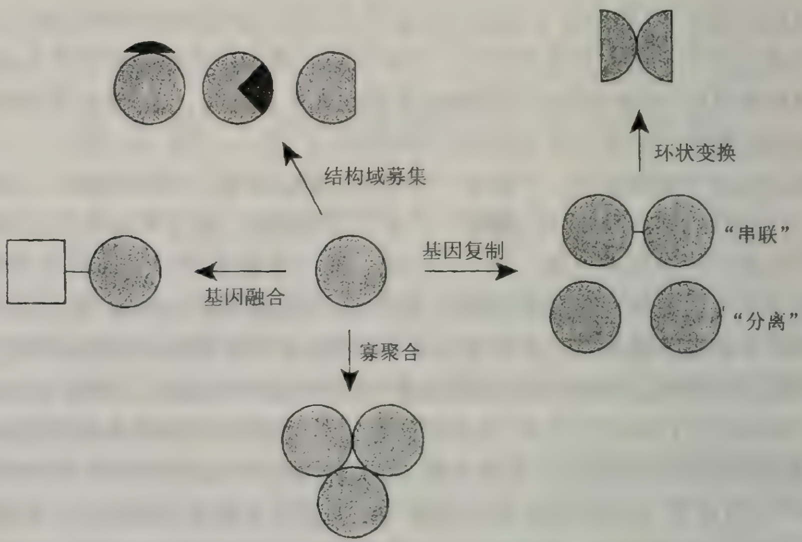


图 5.1 蛋白质进化时主要结构再组织机制的图解

个取新功能。有共同祖先的两种蛋白质功能趋异的范例是组氨酸生物合成途径的 *N*-(磷酸核糖-亚胺甲基)-氨基咪唑-羰酰胺核糖核苷酸异构酶(HisA)和咪唑甘油磷酸合成酶(HisF)的 β/α 桶状结构。两酶共同起源的假说是根据顺序比较提出的,并且由广泛结构相似性鉴定所支持。

5.2.2 串联复制

与局部分开复制的产物相反,基因的串联复制是指基因表达的两个或更多拷贝的产物(图 5.1)。这样的多重性具有以下优点:增加稳定性和新的协同功能;在新的裂缝处形成结合位点;长的重复结构(例如纤维蛋白质)可产生多重结合位点。串联重复可以经历环状变换(circular permutation)和再组织以保持母体顺序和功能的完整性。

经常引用串联复制的例子是胰蛋白酶(天冬氨酸蛋白酶族的一个成员)和胰凝乳蛋白酶(类胰蛋白酶族的代表)。在结构上,两个酶是由两个同源堆积结构域(胰蛋白酶的 β 折叠;胰凝乳蛋白酶的 β 桶)组成,活性部位残基位于结构域的界面上。尽管在个别结构域对之间存在广泛的顺序趋异,但它们总的骨架结构可相互重叠,这表明祖先是单一结构域均二聚体。同样,巨细胞病毒(肝炎病毒族的一个成员)丝氨酸蛋白酶的均二聚体 β 桶结构,似乎是现代丝氨酸酶进化的前体。

串联复制所产生的另一个重要的结构是 $(\beta/\alpha)_8$ 桶,它是自然界普遍存在的型

主(motif),约 10%已知的酶存在此型主。原型$(\beta/\alpha)_8$ 桶是由 8 个股扭转-螺旋-扭转重复组成,通过连续的氢键网络形成内部平行 β 折叠和表面螺旋。前面提及的基因产物 HisA 和 HisF 的多重顺序匹配指出在折叠之内存在内部重复[1]。该资料表明$(\beta\alpha)_8$ 结构域起源于$(\beta\alpha)_4$ 单元的串联复制(图 5.2),这种基因复制假说的正确性已由 HisF 验证。将酶分成两个对称$(\beta\alpha)_4$ 结构域之后并在 *E. coli* 中表达,两个亚基独立地折叠成它的正确构象。当两个$(\beta\alpha)_4$ 单元共表达,自身缔合成功能不均一二聚体,其活性与野生型相似,在某种程序上支持了上述假说。$(\beta\alpha)_4$ 桶本身也可能来自个别的(βα)单元。桶状结构能耐受(βα)单元的重排和取代。

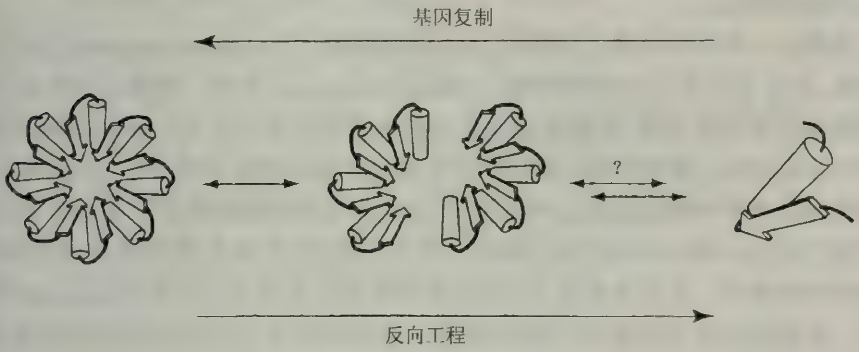


图 5.2　$(\beta/\alpha)_8$ 桶结构的图解说明

高度对称的结构大概来自 βα 片段的多重基因复制。该假说应用逆向工程已检验。将桶分成两半,产生正确的折叠片段,其不均一二聚化恢复母体功能

5.2.3　环状变换

蛋白质的环状变换造成在现有结构框架内 N 和 C 端重新定位。一种环状变换模型指出以前体基因的串联复制起始,在最初末端框架内融合,继之在第一个重复中产生新的起始密码子,在第二个重复产生终止部位。在 DNA 甲基转移酶中观察到按此模型串联复制。

通过环状变换产生 βα 桶折叠结构多样化,最初是在 α-淀粉酶超族各成员的顺序匹配中发现的。后来 *E. coli* 转醛醇酶 B 的结晶学资料分析也指出该酶来自类 1 醛缩酶的环状变换。在上述两种情况下,两个 N 端 βα 重复(加上淀粉酶的第三个亚基的 β 股)移到 C 端,功能没有明显改变。

环状变换进化的优点是不清楚的。然而,体外实验已揭示虽然环状变换的蛋白质最终结构基本上相同,但亚基的重排改变了折叠途径并影响蛋白质构象的稳定性。从工程的观点来看,这样的结构重排的潜在好处是降低结合部位之间的立

体干扰。例如环状变换的白介素 4-假单胞杆菌属外毒素融合物,抗肿瘤活性增加是由于毒素和白介素结合部位之间的立体干扰降低所致。

5.2.4 寡聚合

同源和非同源蛋白质的非共价缔合在酶功能上起重要作用。寡聚合可以调节蛋白质活性或通过二聚体活化转录因子。此外,暴露疏水表面的结构附聚作用增加热力学稳定性,并且可以在蛋白质-蛋白质界面产生穴和沟,其可容纳底物或调节因子结合部位。

寡聚合一种特殊的情况是通过 3D-结构域转换(3D-domain swapping)形成均一二聚体,该术语容易与“结构域转换”(domain swapping)混淆。如图 5.3 所示,后者是结构域募集的同义词,并通过为个别结构域或蛋白质内含子(intein)编码的两个基因片段取代来工程蛋白质,涉及多肽骨架共价键的断裂和重建。相反,3D-结构转换是相同多结构域蛋白质之间结构再组织,没有共价键的断裂。在这个过程中,一个蛋白质的一部分与同一蛋白质的同样单位转换,形成互相盘绕的二聚体或高度有序的寡聚体。天然发生的 3D-结构域转换的几个例子已经由 Schlunegger 等评论[3]。交换结构单元范围从二级结构到完整的结构域,虽然 3D-结构域转换的优点不完全了解,但它可能调节以及增加结构稳定性。

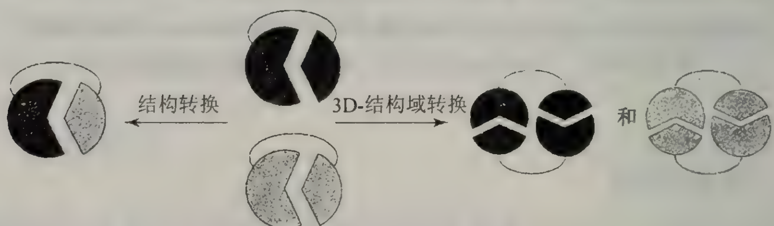


图 5.3 结构转换/募集与 3D-结构域转换之间差别的图解说明

5.2.5 基因融合

完整的融合基因产生共价连接的多功能酶复合物较之个别酶之间非共价寡聚合更具协调性。在生物合成途径中往往发现催化连续步骤的各种酶基因融合可以提供稳定性,调节功能,底物通道,定向协同表达,熵优势和宿主选择优势。

嘧啶和嘌呤生物合成途径中多基因融合已有详细讨论[4],组氨酸和芳香族氨基酸生物合成途径已鉴定[5]。分支酸是色氨酸、酪氨酸及苯丙氨酸生物合成趋异的关键中间体(图 5.4),在 *E. coli* 中,两个双功能复合物催化分支酸经由预苯酸中

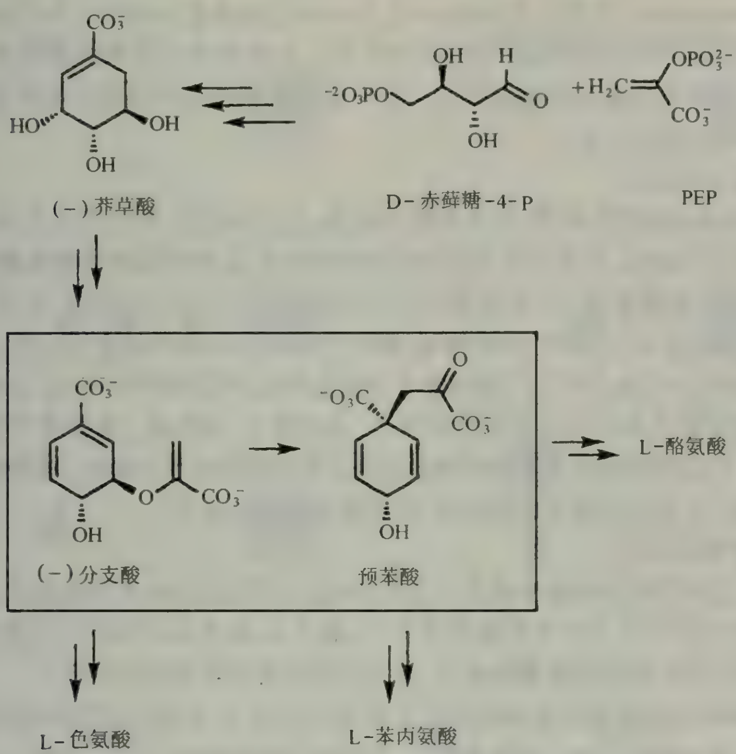


图 5.4 植物和低等有机体中芳香族氨基酸生物合成的莽草酸途径

分支酸[3,3]-σ 移位重排成预苯酸,PEP 是指磷酸烯醇式丙酮酸

间体分别转化为酪氨酸和苯丙氨酸的 α-酮酸前体。酪氨酸前体是分支酸变位酶(chorismate mutase, CM)与预苯酸脱氢酶(CM-PDG)融合合成的,苯丙酮酸是由CM-预苯酸脱水酶复合物(CM-PDR)产生的。两个复合物对终产物反馈抑制敏感。在融合物分成两部分之后,个别的单体仍保持功能(底物通道目的基因融合除外),二聚体形成之后别构控制似乎是驱动力。这些发现与 *N*-乙酰葡萄糖胺-1-磷酸尿苷酰转移酶(催化葡萄糖胺-1-磷酸连续的乙酰化和尿苷酰化酶复合物)分为两部分的实验资料是一致的。两个酶分裂表明,个体酶的动力学、途径总通量和机体的表型均没有明显影响。相反,合理设计的 β-半乳糖苷酶-半乳糖脱氢酶和半乳糖脱氢酶-细菌荧光素酶能够提高底物加工能力。

5.2.6 结构域募集

通过个别蛋白质亚基或结构域再组织产生多功能蛋白质称为“结构域募集”

(domain recruitment),范围从简单的 N 或 C 端融合到多重内部插入和可能的环状变换。最近蛋白质结构数据库(PDB)分析进一步表明作为结构域改组的一种结果,结构重排明显贡献给功能多样性。结构域募集和它们对功能影响的各种模型都是以 βα 桶结构为代表。

(1) 底物特异性

糖水解酶族利用各种寡-和多糖作为底物,它们的活性部位定位在 βα 桶核心结构的 C 端大沟中。族各成员之间的底物特异性是通过附加的结构域调节的,其有效地关闭活性部位和(对着底物)调节活性部位的大小和形状。在 *E. coli* 的 LacZ 中已鉴定出 4 个"附加"的结构域,围绕中心 βα 桶协调地结合在一起。这些附加的结构域完全代表具有不同结构的个别折叠单位:卷饼桶状结构域 1;纤连蛋白Ⅲ折叠(结构域 2 和 4)和 β-夹层结构域 5。在功能上,结构域 1 直接参与形成活性部位口袋,而结构域 2 和 4 作为结构域链节正确定位结构域 1 和 5。结构域 5 的作用不太清楚,它似乎贡献于底物结合,也可能与寡聚合有关。

(2) 调节作用

猪胰 α-淀粉酶(糖水解酶族另一成员)的催化活性是依赖金属辅因子。结合钙离子的部位是位于反平行 β 折叠的界面上,插入 βα 桶核心结构的一个突环区中。在结合钙时,插入稳定结合底物部位,并且间接约束活性部位的构象。

磷酸烯醇式丙酮酸羧化酶的结构分析指出围绕 βα 桶核心结构的活性部位有 40 个螺旋,附加结构域产生天冬氨酸(一种酶的别构抑制剂)的结合部位。该酶的高度保守的 C 端延伸大概是寡聚合的关键。

(3) 转移电子

Sacchromyces cerevisiae 黄细胞色素 b2(依赖 FMN 氧化还原酶超族的一个成员)催化乳酸到丙酮酸,继之结合黄素使电子转移到细胞色素 C 上。该酶与超族其他成员的区别是结合血红素结构域的 N 端与 βα 桶状结构融合形成基本的活性部位。在这个过程中,不是把还原黄素氢醌的电子抛到分子氧上,而是在分子内部转移到结合血红素的结构域并经分子间步骤转移到细胞色素 C 上。

5.2.7　外显子改组

涉及血液凝集和血纤维蛋白溶解作用的几个蛋白酶的结构分析揭示出多样性、重复性、模块组装性质(图 5.5)。这些模块代表独立的折叠单位,它们的协同作用最终贡献给蛋白质的功能和特异性。在遗传水平上,这些模块由外显子编码。在模块蛋白质进化过程中,新基因是通过外显子复制、缺失和重排产生的。外显子重组实际发生在插入的内含子顺序(内含子重组)[6]。

蛋白质进化,特别是多细胞发育都涉及到了模块组装所产生的各种蛋白质的

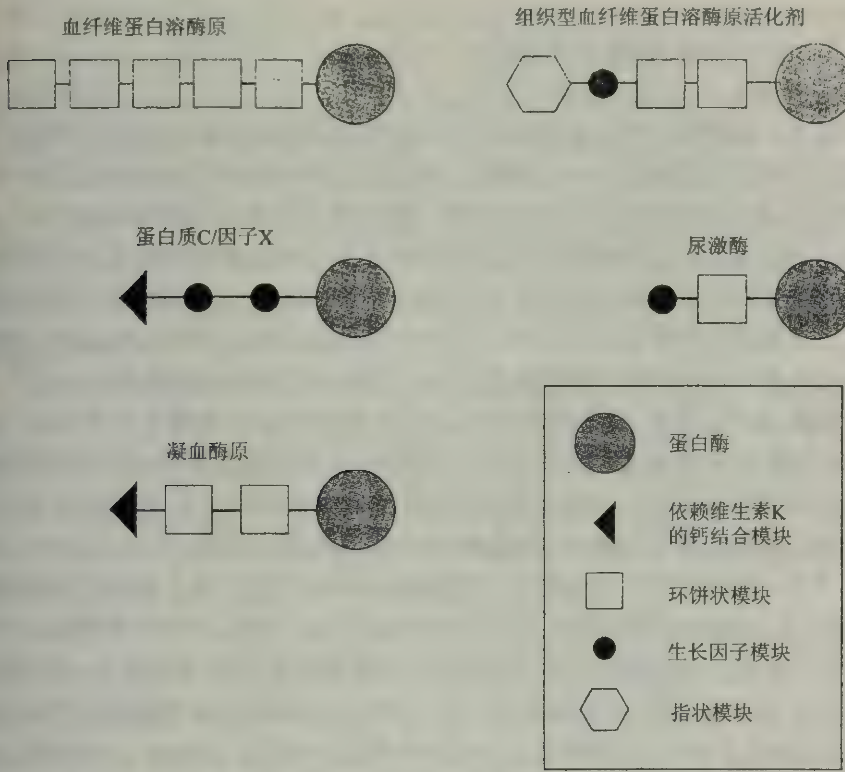


图 5.5　血液凝聚和血纤维蛋白溶解的蛋白酶模块进化的图解说明
图中指出融合到共同的蛋白酶单位的模块类型和数量，产生高度特异的水解酶族

加工、外显子改组涉及血液凝集、血纤维蛋白溶解和补体活化的调节蛋白质、细胞基质的大多数成分、细胞黏连蛋白质和受体蛋白质的构建。

5.3　定向分子进化

5.3.1　组合定向进化

过去 10 年，科学家发展了一系列模拟自然进化机制和过程的方法进化现有的酶(见第 2 章)，组合进化是特别活跃的研究领域。组合进化促进了蛋白质结构与功能关系的阐明、功能与进化的了解，同时也提供进化酶。

组合进化的主要方法之一是随机-定位突变，又称为随机广泛突变(random and extensive mutagenesis，REM)。

5.3.1.1　L-天冬氨酸酶基因 REM

现以天冬氨酸酶为例介绍 REM[7,8]。REM 方法的基本原理见第 2 章图 2.9，它是通过限制某特定碱基(如 A)的浓度，使引物沿模板延伸到与该碱基配对的碱基前(如 T)，由此产生长短不一的分子群。酶法掺入另外 3 个错配碱基(如 C，G，T)引起点突变，合成片段继续延伸，经连接酶连接、转化、尿嘧啶模板法筛选得突变株。该法有效地避免野生型再生，产生所有可能的突变又不需要蛋白质空间结构信息指导便获得所希望的突变体，还可灵活控制基因突变的随机和定位性(即突变位点的多寡)，从而控制突变库的大小，也可控制突变碱基的准确变化。总之，该法具有简便、灵活、高效、耗资低等特点。由突变库筛选 2 株酶活力约为 *E. coli* AS1.881 菌株 10 倍和 15 倍的工程菌。在此基础上，根据化学修饰和顺序同源性分析，用 Arg 取代 Lys-125(K125R)，在结构没有明显改变的情况下，功能却发生了变化。K125R 突变酶最适 pH8.5(天然酶 pH8.0)，pH 稳定范围 7.0～9.0(天然酶 pH7.0～8.0)，热稳定性好于天然酶，突变酶活力提高 5 倍。为了保证工程菌能稳定地遗传，在工程菌的质粒中插入来自 pSC101 的 par 片段，从而获得酶活力较高且稳定的工程菌。该工程菌培养 100 代后含质粒菌株数仍在 90% 以上，并且培养温度升高到 40℃时，仍不受明显影响。将 K125R 突变菌株用琼脂凝胶包埋法进行固定，固定菌量为 95%，表观酶活力为 82 000 单位，半衰期为 184d。实验表明，1kg 固定化细胞，每 24h 约得 L-天冬氨酸 30kg，产品质量达 1987 年日本药局方(第九版)规定的标准。

5.3.1.2　α-天冬氨酰二肽酶基因 REM

下面我们以 α-天冬氨酰二肽酶为例来介绍 REM 法[9,10]，由于 α-天冬氨酰二肽酶较为特殊，目前，人们只发现另外一种酶与其有同源性，所以我们无法利用现有的蛋白质数据库通过计算机预测出该酶的三级结构。幸运的是，无活性 α-天冬氨酰二肽酶突变体 I 的基因(*pepE I*)顺序分析表明，在读码框内发生两处碱基替代，致使 137 位脯氨酸变为丝氨酸，120 位的氨基酸由原来的丝氨酸变为苏氨酸。突变酶无活性可能是由于 Pro137 变为 Ser137 的结果。我们分别对天然酶、突变酶(P137S，S120T)进行了二级结构分析，二级结构未见明显改变。由于 Pro 为亚氨基酸(N 上只有 1 个氢原子)，到目前为止，还未发现其存在于活性部位，因此我们推测，双突变体没有活性可能与 Ser120 有关。

虽然 Ala 不是取代任何氨基酸的最适候选者，但由于 Ala 是一个比较“随和”的氨基酸，可以出现在任何二级结构中，以它替换其他氨基酸时，蛋白质分子构象

一般不会发生太大的改变；另外 Ala 没有催化活性，如果某一氨基酸被突变成 Ala 后活性没有太大的改变，那么这个残基不是酶活性必需的。所以我们的策略是用丙氨酸扫描(alanine-scanning)法，先用 Ala 分别替换值得怀疑的氨基酸，排除一些突变后相对来说不影响酶活性的残基，再对表现出重要性的残基进一步研究。

对 Ser120 的定位诱变实验结果表明，α-天冬氨酰二肽酶 Ser120 变为 Ala 或 Thr 后，突变酶丧失活性，说明 Ser120 是酶活性部位的氨基酸残基，考虑到 Ser 与 α-天冬氨酰二肽酶没有特殊的匹配关系，如果 Ser120 是结合部位的话，就不能解释 α-天冬氨酰二肽酶高度的底物专一性，以及突变后的活力完全丧失。因此可以得出结论，该酶应属于丝氨酸蛋白酶类。

根据化学修饰的结果，精氨酸和组氨基对酶的活性是必需的。精氨酸不太可能直接参与催化功能，而且该酶对 Asp 带负电荷的侧链高度的专一性，暗示了精氨酸残基很可能处于底物的结合部位，所以我们选择了组氨酸作为突变的对象，来研究哪个组氨酸残基位于催化部位。通过二级结构预测，酶分子中的 5 个组氨酸中有两个(His19、His227)位于 α 螺旋中，所以我们优先考虑了 His68、His157、His166。定位诱变的结果表明，His157 是活性部位的氨基酸残基。

研究表明，α-天冬氨酰二肽酶属于丝氨酸蛋白酶类，其催化部位的氨基酸残基包括 Ser120、His157、Glu192。其催化二肽 L-Asp-X 的水解机制，据推测应类似于其他丝氨酸蛋白酶。首先酶与底物结合，形成一个非共价的酶-底物复合体，Ser120 的羟基氧对底物肽键上的羰基碳进行亲核攻击，形成四面体中间体；His157 和 Glu，起广义酸碱催化作用，催化四面体分解，形成酰化酶中间体，同时放出胺。水作为亲核试剂攻击酰化酶中间体，形成酶-底物复合物，脱酰化作用生成产物 L-Asp 和 X。

α-天冬氨酰二肽酶的 3D 结构已解析(图 5.6)[9]。它是由两个混合的 β 折叠形成“V”形，两侧有 6 个 α 螺旋。活性部位含有 Ser-His-Glu 催化三元组(图 5.7)[9]。活性部位 Ser120 位于较大的 β 折叠上，His157 位于较小的折叠上，Glu192 位于两个 β 折叠之间短的交叉片段上。它是第一个催化三元组有 Glu 的肽键水解酶。

α-天冬氨酰二肽酶的定向进化获得一进化酶，酶活性比野生型提高 47 倍，并且稳定性明显增加[10]。顺序分析表明有 3 个氨基酸取代(Asn89His, Gln153Glu, Leu205Arg)。从 3D 结构模型上看，Asn89His 非常接近活性部位的氨基酸残基；Gln153Glu 距离 His157 仅有 4 个氨基酸残基，而 Leu205Arg 远离活性部位[图 5.8(a)][10]。残基 89 定位于连接第四个 α 螺旋和较大 β 折叠的第一股之间的突环上。有意义的是残基 89 是 GGXT 型主(残基 87 ~ 90，X 是任一氨基酸)中非保守氨基酸，它是活性部位构象的关键，将明显影响活性而不破坏二级结构[图 5.8(b)]。His157 的 Nδ1 与 Glu192 的 β-羧基的氢键增加了 His 的 Nε2 的亲核性，它通过氢键同 Ser120 的羟基相互作用。结果 Ser120 侧链的氧原子成为强有力的亲核试剂，进

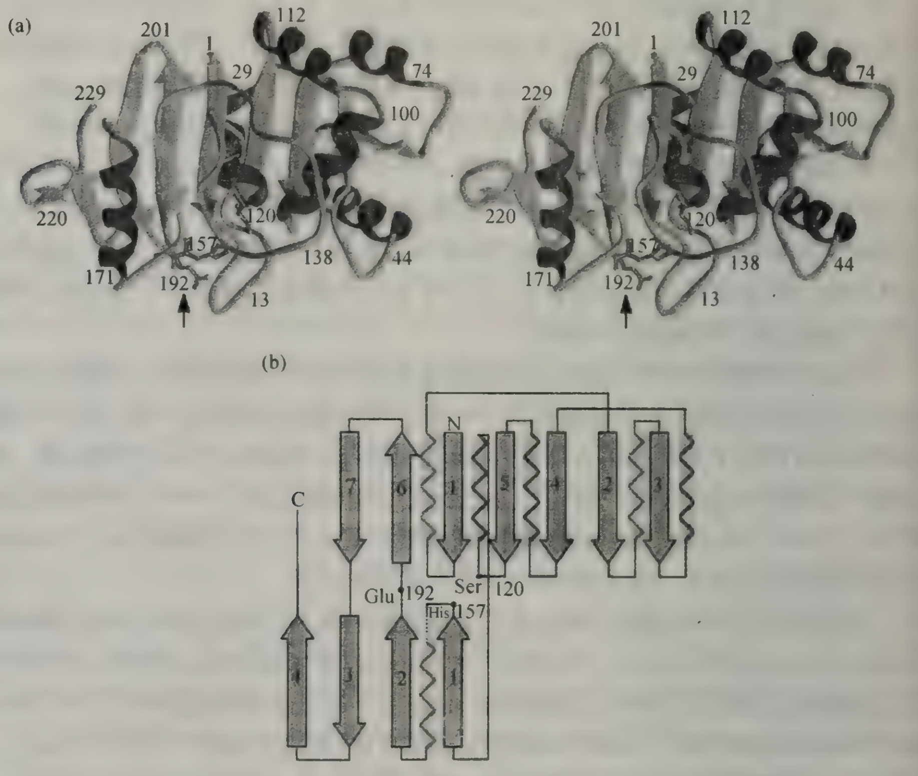


图 5.6　α-天冬氨酰二肽酶的结构

(a)带状图解,箭头表示催化三元组残基的侧链;(b)拓扑图解

攻底物羰基碳原子,形成四面体中间体。His89 大概有助于稳定过渡态的负碳离子中间体(图 5.9),提高进化酶的 K_{cat}并且非常接近底物结合口袋。Asn89His 可以引起底物的结合和定位的微妙变化,并导致酶与底物亲和性的提高。Gln153 定位在较小 β 折叠的第一股末端,离活性部位 His 仅 4 个残基(在进化酶中却是 Glu153)。这个取代可能微妙地调整 His157 的 Nε2 更有利与 Ser120 的羟基形成氢键。然而进化酶的 Glu153 的羧基大概同 Arg131 的胍基形成内部离子对[图 5.8(c)],提高进化酶的热稳定性。Leu205Arg 的作用难以说明。

我们在 L-天冬氨酸酶进化酶四聚体和 α-天冬氨酰二肽酶进化酶(单体)基础上,通过体外选择获得新的杂合酶,称之为进化型杂合酶(evolved hybrid enzyme,EHE)。EHE 具有 340% 的 α-天冬氨酰二肽酶的活性和 87% 的 L-天冬氨酸酶活性。在酶浓度为 10^{-7}mol/L 时,EHE 的活性可逆性略高于天然酶。该结果表明由进化酶构建杂合酶可能是一种有效的途径。

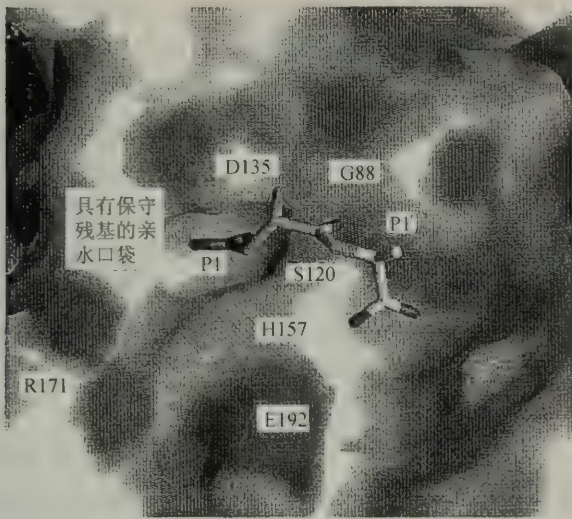


图 5.7　α-天冬氨酰二肽酶的活性部位

Asp-Ala 二肽与活性部位结合。底物的氨基同 Asp135 相互作用。P1 侧链是位于具有保守残基的疏水口袋中，P1'侧链朝向溶剂。只有在羰基原子和活性 Ser 羟基之间紧密的接触时，底物可以进一步运动到活性部位

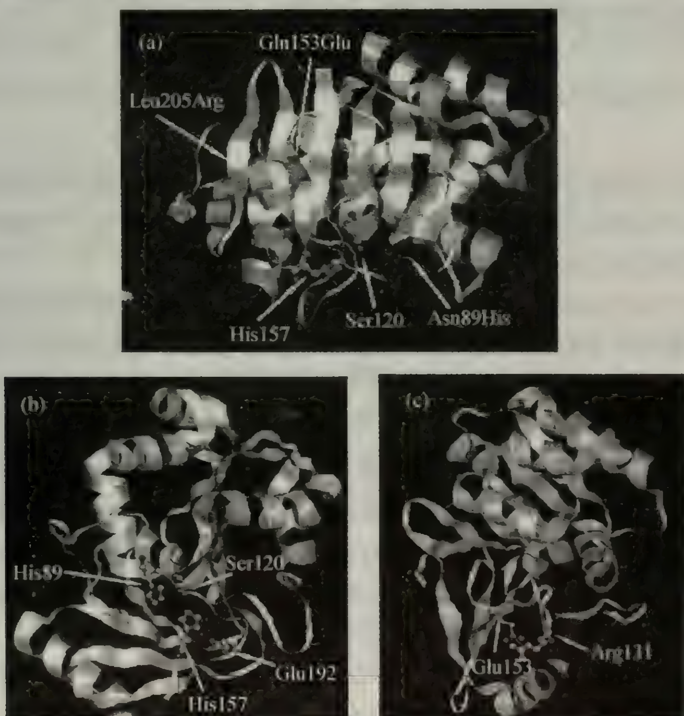


图 5.8　野生型 α-天冬氨酰二肽酶突变体 PepEM3074 的结构模型

(a)突变位置(Asn89His、Gln153Glu 和 Leu205Arg)；(b)His89 的空间位置和活性部位；(c)Glu153 和 Arg131 的位置

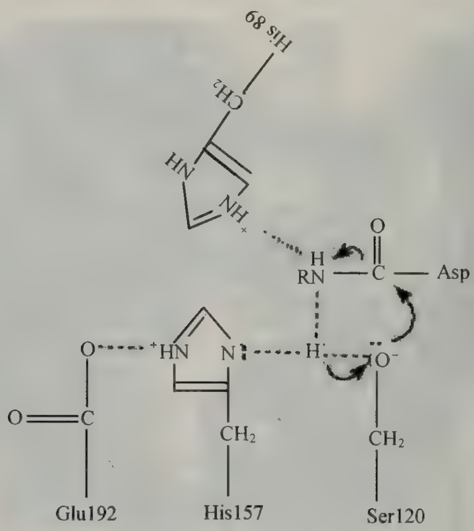


图 5.9　α-天冬氨酰二肽酶的进化酶 His89 如何加速活性部位反应速率的假设说明

5.3.2　模块定向进化

5.3.2.1　模块酶[11]

模块酶(modular enzyme)是一个多组分系统,其中单独的组分可以与其相关系统中功能不同的模块进行交换。现在,酶的模块性质已经成为分子酶学工程的强有力的概念和技术基础。

自然界中的模块酶并不普遍,然而它对化学家的潜在的利用价值是巨大的。由模块酶的数量和多样性出发,使人们能够根据需要设计出各种各样的新酶。

1. 模块酶的概念

所谓模块酶就是具有模块性质的酶。对一个有模块性质的酶而言,它最低必须有一个模块结构。典型的是多结构域或寡聚酶中的各亚基。模块性质需要满足两个标准:第一,不同的结构域有不同的功能;第二,这些结构域可以重组产生嵌合体。目前只有少数酶族可以满足这两个标准。但是迅速扩展的顺序和结构数据库将会增加模块酶族的目录。模块酶类似于模板,例如,催化核糖体蛋白质合成的装置。在大多数情况下,虽然单独结构域或亚基的性质可以重组成嵌合体,但形成的嵌合体与其母体相比较存在着动力学缺陷。

自然界模块酶有三种情况:①酶的催化和底物特异性是分开的;②在多底物酶

中单独底物的结合部位是模块；③按程序催化的代谢途径的多酶系统。下面，我们将选择实例说明属于上述情况的每一种酶的性质。我们也讨论在每种情况工程酶的机会，以及它们潜在的应用价值。作为一种参考框架，我们指出原型模块酶-核糖体蛋白质生物合成装置表现出所有三种类型模块性质。催化剂（核糖体）可以容易同支配底物特异性的单元（mRNA 模板）分开。对于肽键形成反应来讲，两个乙酰化转移 RNA 底物与核糖体的不同位点（A 位和 P 位）结合。这个模块系统所催化的多步途径可以通过密码子去除，加入或取代而程序化。关于依赖模板的 DNA 和 RNA 聚合酶也是如此。这些模块特征是构建实效生物催化剂惟一成功的基础。因此，更好地了解模块酶的化学和生物学将对生物催化剂的开发具有重要的意义。

2. 模块酶的类型

(1) 催化与分子识别分离

如果一个酶的化学反应性可以以模块方式与其底物识别分离，那么它就有可能产生特殊的和一般的催化剂（或一族催化剂）。这是难以办到的，原因是底物结合和催化活性是连在一起。结合时各种相互作用来调整底物相对于酶活性部位的残基的方位，并且优先稳定过渡态。对于部分底物结合和催化未偶联（模块方式）的酶的开发，至少有三种不同的策略。它们包括不同结构域的催化和识别功能分离；可变换的底物结合口袋的进化；一般底物中模块的化学特征（化学模块）的识别。

*Fok*I 限制性内切核酸酶是由两个独立的折叠结构域组成，它可通过有限水解分开（图 5.10）。第一个结构域结合一个 GGATG DNA 识别型主，第二个结构域催化 DNA 股分裂。将 *Fok*I 分裂的结构域与不相关的转录因子的 DNA 结合结构域融合，从而产生具有新的分裂特异性的嵌合体。

降解植物细胞壁的各种细菌水解酶也表现出底物识别和催化之间结构域部分分离。这些水解酶（包括木聚糖酶和纤维素酶）除了一系列结合碳水化合物的结构域（CBDs）外，还含有一个或两个糖苷酶的催化结构域。催化结构域可以水解可溶性寡聚糖，但是它们降解植物细胞壁中结晶的糖聚合物需要 CBD 结构域。CBD 的作用是增加酶-底物接触。结合纤维素和木聚糖的结构域与异源的糖苷酶结构域融合能够代谢结晶的底物。

在不同基因组中共价结构域模式分析可以发现具有独立催化和结合结构域的其他模块酶。但是这类酶的某些机械限制最重要的是通过识别结构域的结合不能确定反应底物的方位。事实上，结构域之间的链节必须是柔性的，所以，通过识别结构域的结合不能妨碍底物在活性部位的正确定位。结合底物的有效浓度通过结构域间的链节更接近催化结构域，从而决定酶的选择性。

除了基于结构域的分离外，催化作用也可通过模块结合口袋交换（作为活性部



图 5.10 *Fok*I 限制性内切核酸酶的催化和分子识别结构域的分离
C 端分裂 DNA 的结构域(Ⅰ);N 端结合 DNA 的结构域(Ⅱ);DNA 双螺旋(Ⅲ)

位定位在同一结构域)与底物识别分离。四膜虫组Ⅰ核酶就是这类模块性质的酶。该核酶内部的指导顺序(internal guide sequence, IGS)同 RNA 底物形成复合体,给分裂的底物以定位。底物和 IGS 之间的 Watson-Crick 碱基配对控制底物的识别,IGS 的碱基取代可以产生对 RNA 顺序特异的核酶。同样,激酶的 Src 族提供酶同非多聚物底物的结合口袋模块性质又一个实例。在 Src 或 Fyn 激酶的核苷酸结合口袋中 Ile→Gly,这个突变酶可以有效地利用 N^6-苄基 ATP 作底物。具有转换底物结合位点的嵌合酶往往表现出新的底物亲和力,但这种亲和力与底物转化为产物的能力无关。酶族的基因改组可能是把结合口袋模块性质普遍化的一种有效工具。

有些酶的催化和分子识别通过识别其底物中特殊的化学特征(化学模块)而模块化。例如,脂肪酶(直链碳水化合物)、青霉素 G 酰化酶(苯乙酰基)、丁酰基胆碱酯酶(胆碱)、邻苯二甲酰胺酶(邻苯二甲酰胺)和芳香基乙酰胺酶(乙酰替苯胺)。这些酶正在用于处理复杂分子的保护基。例如,*Serratia marcescens* 脂肪酶已用于生产钙拮抗剂的一种合成中间体——diltiazem,年产规模 50t。

到目前为止,已经发现了一些化学模块酶。青霉素 G 酰化酶(图 5.11)和几个脂肪酶的结晶结构指出了它们的作用模型。两者的结构揭示活性部位指向溶剂。化学模块酶是环境污染的"清洁工",例如植物,根瘤菌和土壤微生物的芳香基乙酰胺酶降解各种 *N*-乙酰苯胺衍生的农药(例如 Propanil)。丁酰基胆碱酯酶存在于人的肝脏和血浆中,它清除几种胆碱酯,并解除大量的药物毒性,包括可卡因。涉及化学战的各种酶可以在耐受各种底物选择压力下进化,涉及这些过程的酶可以预

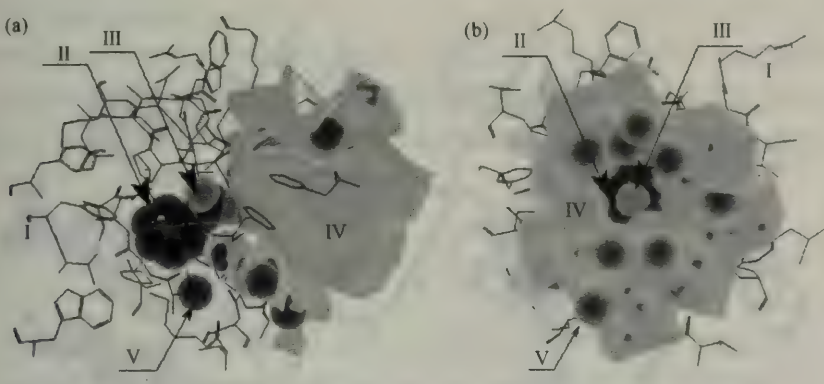


图5.11　青霉素G酰化酶识别底物中化学模块(苯乙酰基)

活性部位的残基(I),结合苯乙酸的碳(II)和氧原子(III),酶的部分溶剂可及表面(IV),有序水分子(V)。(a) 苯乙酰基部位被酶包围,酸的末端氧原子指向溶剂,青霉素酰化酶的各种胺和乙醇底物通过末端原子系在苯乙酰保护基上。(b) 同一活性部位围绕垂直轴旋转90°。苯乙酸末端氧原子通过溶剂可及表面可以看见口袋。该图是用INSIGHTII蛋白质数据库IPNL坐标产生的

期是发现新酶的源泉。

(2) 多底物酶中分子识别特征的分离

有一些酶催化两个或更多的底物反应。在第一个产物形成之前,所有底物与酶结合的反应称为顺序(sequential),而释放一个或更多产物的反应称为乒乓。顺序酶的活性部位可以随机或强制程序结合底物。在乒乓酶或强制程序中,两者的底物结合口袋往往重叠或彼此相互作用。但是在随机结合时,双底物酶的结合口袋一般是分离的,在某种情况下,它们甚至可以位于多肽骨架的不同结构域。这样的模块酶的一个典型例子是蛋氨酸合成酶,其底物-高半胱氨酸、甲基四氢叶酸、钴胺素和腺苷基蛋氨酸,它们结合在蛋白质的不同区域。多底物酶的分子识别的模块性质代表了分子酶学工程的一个崭新的起点。

顺序和结构分析指出,许多进化上相关的双底物酶族都有模块分子识别特征。利用核苷二磷酸(NDP)-糖的酶,例如糖苷基转移酶,似乎具有这种性质。NDP-糖苷基转移酶属于大的酶族,它们对糖和糖苷配基具有转变的特异性。最近这个族的原型结晶结构揭示出两瓣结构(图5.12)。一个结构域与NDP-糖结合,而另一结构域与糖苷配基结合。两个结构域由伸展的突环分离,并且活性部位残基位于两个结构域的界面上。顺序比较指出,这个酶的超族大多数成员保持双结构域结构并且活性部位残基定位在两个结构域的界面上。双结构域的分子识别特征不仅独立地进化,而且往往交换而不破坏活性部位的几何形状。这样,单独结构域的宽松的特异性与该酶族的生物合成的重要性一起成为结构域改组的特别诱人的目标。

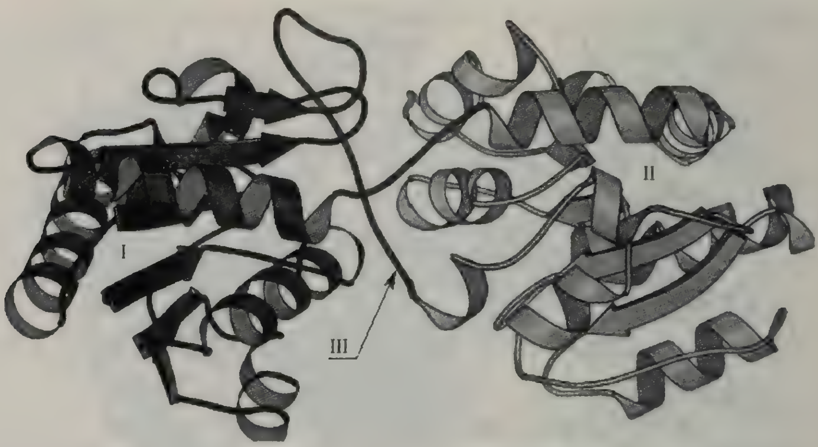


图 5.12　模块多底物酶的分子识别特征

双结构域 MurG 糖基转移酶。NDP-糖和糖苷配基结合结构域分为以(I)和(II)表示。连接这两个结构域的两个突环以(III)表示。活性部位残基位于两个结构域的界面上

(3) 多酶体系中间体通道

PKSs 和 NRPSs 体系还说明模块催化另一原理,这对 one-pot 生物催化过程的发展将具有重要意义。它们代表了一个高度进化的模块策略——由一个活性部位到另一个活性部位中间体通道策略。通道(channeling)可以定义为在没有同整体体液平衡的情况下,代谢顺序中连续的各种酶之间中间体的直接转移。这在色氨酸合成酶和某些氨基转移酶等已得到证实。结晶学已鉴定出通道宽 10～100Å。约束吲哚和氨由一个反应中心向下一个反应中心扩散。对于需要在活性部位之间转移的活泼中间体(例如氨),通道可能是有效的,但受体酶对同源底物的特异性也降低。

多酶组装,例如,α-酮-脱氢酶、脂肪酸合成酶、PKSs 和 NRPSs 有"摇摆臂"(swinging arm)(在脱氢酶中有酯酰胺,在脂肪酸合成酶、PKSs 和 NRPSs 中有磷泛酰巯基乙胺)。这些摇摆臂是柔性的,长约 10～15Å,通道共价结合在连续的活性部位之间的中间体上。两个例子说明了可能的机制,摇摆可以同其他的结构特征组合产生代谢途径模块。第一,为了将辅助的催化结构域嫁接到核心 PKS 模块上。这不仅表明模块的可塑性,而且也指出摇摆臂的存在以提供嫁接新结构域的可行性。第二,选择蛋白质-蛋白质相互作用组合,调整摇摆臂的化学结构,可以促进天然的和非天然的中间体的转移。由于摇摆臂的需要,有三个问题值得注意:①这些机制似乎局限于共价催化系统;②选择链节以指导所设计的催化模块之间的中间体;③不是所有的蛋白质—蛋白质相互作用都是模板。例如,在 PKS 模块中,如果供体载体结构域和受体缩合酶之间的相互作用对链转移是重要的话,那么

这些性质将随模块的超结构而变化。

分子生物学和细胞生物学反复地证明,生物大分子的模块性质可以进化和工程。根据模块性质的结构和机制的了解将会构建出各种各样的模块酶,成为分子酶学工程的重要领域。随着顺序和结构数据库的不断增长,模块酶将迅速地成为工业上有效的生物催化剂。

5.3.2.2　人工模块的随机化

L-天冬氨酸酶(EC4.3.1.1)可逆催化 L-天冬氨酸转为富马酸和 NH_4^+。*E. coli* W 的 L-天冬氨酸酶是由 4 个相同亚基组成,并呈点对称 $P2_12_12_1$ 排列,单亚基无催化功能,每个亚基由 3 个结构域组成(D1,D2 和 D3)呈"S"形[图 5.10(a)]。中央结构域(D2)含有一半以上的氨基酸残基,它是天冬氨酸酶-富马酸酶族最保守的结构域。D2 的核心是由 5 个长螺旋构成,其轻微弯曲并近于彼此平行,它对于保持活性构象至关重要。由于亚基间相互作用,来自 4 个亚基的 D2 五个螺旋束形成一个稳定的 20 螺旋簇,从而构成四聚体[图 5.13(b)]。

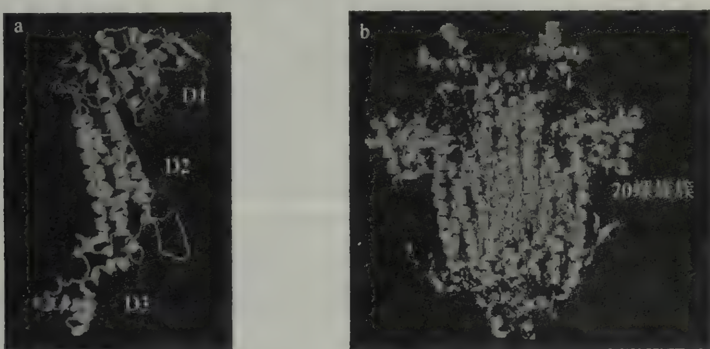


图 5.13　L-天冬氨酸酶亚基(a)及其四聚体(b)的 Molscript 图解
螺旋,股和突环分别以带形,箭状和线表示,图中标出四聚体中的 20 螺旋簇

L-天冬氨酸酶是一个重要的工业用酶,主要用于合成 L-天冬氨酸。然而由于该酶不肽稳定,实际应用受到限制。作为四聚体蛋白质,L-天冬氨酸酶在工业生产条件下容易解离,伴随其活性降低。在 0.4mol/L 盐酸胍存在下,L-天冬氨酸酶可逆地解离成二聚体,其活力为天然酶的 45%。荧光和 CD 研究表明,二聚体的二级和三级结构仅发生小的改变。这些结果表明,二聚体具有活性部位的空间结构,但不像四聚体那样完美。很可能,两个亚基的反向排列决定在空间位置上 D1 和 D3 之间彼此接触,在来自不同亚基的 D1 和 D3 之间的界面上形成活性部位。假若如此,通过模拟二聚体的部分空间结构——中央 10 螺旋束和上述的"活性部位",我

们设计了 L-天冬氨酸酶的突变体,它是由两个模块(D1D2 和 D2D3)通过一个六肽突环连接形成一个单体,来自重复 D2 的两个五螺旋束彼此相互作用恰像来自不同亚基,如果突环采取适当的构象,180°扭转,允许 D2D3(或 D1D2)回折,并取代其他的亚基形成 10 螺旋束,则类似于二聚体(图 5.14)。如果 D1 和 D3 域间接触则形成一个潜在的活性部位。为此,最好的方法是 D1D2 和 D2D3 之间的插入顺序随机化,继之根据 L-天冬氨酸酶的活性体外选择。我们由组合库中得到了一个高活性的突变体(相对活性是天然酶的 21.3%)。通过大小排阻色谱法(size-exclusion chromatography)测定分子质量为 81kDa,的确是一单体。这个单体较之天然酶具有一些更适于工业应用的特性:活性可逆性高于天然酶,单体蛋白质避免解离成非活性形式,即稳定;如果作为医药,它具有较低的免疫原性[12]。所描述的方法可以作为研究如何将寡核体变为单体的模型系统,也可以加深了解寡聚酶的进化起源。

5.3.2.3 天然模块的随机化

多功能酶系统中一个最活跃的研究内容是通过聚酮合酶和非核糖体肽合酶合成各种有机化合物[13~15]。一般,这些酶是由多亚基表征,每个亚基是由不同酶活性的结构域组成[图 5.15(a)]。通过这些"巨酶"(meta-enzyme)合成的天然化合物包括各种重要的抗生素、抗真菌、抗肿瘤和降低胆固醇化合物以及免疫抑制剂等。

根据"巨酶"的底物特异性和产物合成的方式可将它们分为 3 种类型。利用乙酰辅酶 A(CoA)单体的两个已知类型的聚酮合酶(PKS)(类型 I 和 II),而非核糖体肽合酶(NRPS)用氨基酸和它们的类似物作为底物。类型 I PKS 和 NRPS 通过模块组装,而类型 II 通过重复单体单元组装。

PKS 和 NPRS 的模块设计产生杂合酶,可以合成天然和非天然代谢物。在聚酮合成中有 4 个自由度:链长、ACP 结构域的选择(底物)、β-碳的修饰和立体化学和形成高度多样性的产物库。同样,在 NRPS 中,链长、底物的类型和立体化学,以及它们后合成衍生化水平可以产生新的或改进性质大量化合物。

最初合理工程 PKS 和 NRPS 集注在改变生物合成途径,结构域的缺失,插入和取代[图 5.15(b)]。虽然这些实验提供了关于结构域和模块的结构与功能方面有价值的信息,但功能杂合酶的产生是困难的。相反,取整个模块并按酶复合物中前后关系排列它们,这是产生功能杂合酶比较现实的路线,共价及非共价连接模块之间的链节区是目标。通过重建之间-多肽链节顺序或之内-多肽链节顺序,转换 6-脱氧红霉内酯 B 合成酶(DEBS)中的模块位置,并引入其他的 PKS 模块,产生一系列杂合酶。

除了完整模块转换外,为了构建各种化合物库还对每个模块内蛋白质随机化。所设计的类型 I PKS 的模块对于构建组合库是可行的。通过盒式取代 DEBS 的结

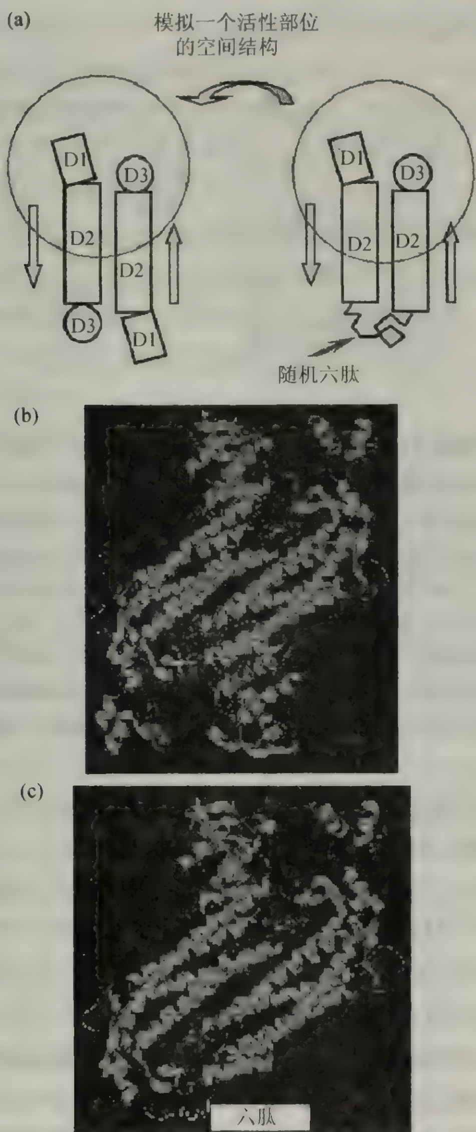
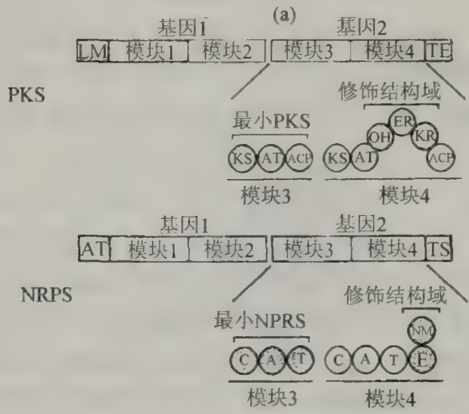


图 5.14　(a) 二聚体和单体 L-天冬氨酸酶的示意图;(b) 二聚体的计算机模型;(c) 结构域重排的 L-天冬氨酸酶计算机模型

构域构建组合库及一种相对简单且优秀的多质粒途径也已描述。DEBS 的 3 个基因被克隆在分离的质粒载体上,继之把突变引入每个基因使其多样化,将它们转化到异源的宿主中,根据 3 个载体随机组合产生相关的聚酮化合物小库。虽然该结



(b)

	PKS	RPS
取代的结构域	173,174	–
合理结构域交换	157–160	151,162,162
合理模块交换	144,163–165	152
组合结构域/模块交换	147,1498	–
结构域修饰	–	153
结构域/模块缺失		149,166–168
混合PKS/NRPS	169–172	

图 5.15　聚酮合酶(PKS)和非核糖体肽合酶(NRPS)的蛋白质工程

(a) 在两个系统中,完整的合酶是由单独的各种基因编码的多亚基组成的。亚基本身分为各种模块,每个催化一个酰基构件(对 PKS 而言)或氨基酸(对 NRPS 而言)的加成。PKS 最小模块是由 3 个结构域组成:一个酮基合成酶(KS)、一个酰基转移酶(AT)和一个酰基载体蛋白(ACP)。此外,模块可以含有 3 个修饰的结构域以衍生聚酮链的 β-碳:一个酮基还原酶(KR)、一个脱氢酶(DH)和一个烯醇还原酶(ER)。整个复合物两侧有负载模块(LM)和终止模块(TE)。NRPS 最小模块是由一个腺苷酸化(A),一个硫代化(T)和一个缩合(C)结构域组成。肽链的附加修饰可以通过差向异构化引入(E)或 N-甲基化(NM)结构域。NRPS 中负载的模块是由腺苷酸化/硫代化(AT)结构域组成,而合成是通过硫酯酶(TS)完成。(b) PKS 和 NRPS 的蛋白质工程

果是令人激动的,但是该方法是受单一宿主细胞可以存在的各种质粒数严重限制,三重转化的效率是低的,质粒稳定性是主要问题。

重复(类型 II)PKS 工程更困难,这是因为它们在天然产物的生物合成中重复使用单一模块。因此,只用该类型的 PKS 所产生的聚酮化合物的组合库不能由遗传编码多样性构建,而是通过最小 PKS 组装。最小 PKS 通过去除所有的附属蛋白质[例如环化酶、酮基还原酶及各种芳香族酶(aromatase)],把功能酶复合物减少到 3 个主要的结构域:一个酮合成酶、一个链长因子和一个酰基载体蛋白质。whiE 最小变体是单一酶复合物,它可以合成 24 种聚酮化合物;*Streptomyces maritimus* PKS 应用同样的自然机制产生一套天然色素。相反,指导特殊代谢产物生物合成的自然机制是通过附属的蛋白质(例如天然 whiE 系统中的环化酶)完成的。通过与最小 PKS 复合物非共价结合,附属的蛋白质稳定产物-酶中间体,保证形成全长的产物和正确的环化。这样一种系统对于产生聚酮化合物库是有价值的。库组成的再现性可能是个问题,这是因为影响多样性产生的各种因子是未知的,而且对库中目标产物的鉴定没有直接的方法。因此,需要附加的合理设计来构建专一产生所希望化合物的杂合体。

非核糖体肽合酶是多酶复合物族成员的代表。它们代表一种重要的和通用的巨合成酶类,涉及肽基次级代谢产物(包括青霉素、万古霉素、博来霉素和环孢霉素A)广谱生物合成。这些化合物的生物合成不仅仅利用天然氨基酸,而且也广为利用 D-构型和 *N*-甲基化的氨基酸以及羟基酸。肽骨架结构的变化形成线性的,环状的和分支的环状分子,并且这些结构可以进行酰化,杂环化,还原和糖基化修饰。与类型 I PKS 的模块设计非常相似,非核糖体肽合酶是由共价或非共价连接多重模块线性排列而成[图 5.15(a)]。个别的模块负责特殊的氨基酸移入到生长的肽链中。模块本身是由 3 个结构域组成的:一个腺苷化结构域(A 结构域;活化进入的氨基酸)、一个硫代化结构域(T 结构域;肽基载体蛋白)和一个缩合结构域(C 结构域;促进肽基部位和活化氨基酸之间的反应)。像 PKS,合理设计实验已集中在一个模块内的个别结构域。

通过 NRPS 产生肽基库的组合途径尚未见报道。考虑 NRPS 与模块 PKS 总体设计非常相似,人们可以想像应用相似的方法产生肽基库。NRPS 的底物-识别 A-结构域的定位诱变结果产生底物特异性的改变或松弛,可以用于产生异源产物库。

总之,PKS 和 NRPS 的结构与功能的研究为级代谢产物的生物合成提供了重要的基础信息。

非核糖体肽合酶是模块“巨酶”,其应用于多个载体硫代模板机制催化组装大量的生物活性肽。各种模块包含特殊的结构域,其催化与底物活化修饰和缩合有关的特殊反应。一套双模块杂合 NRPS 的设计和构建已有报道;通过引入腺苷酸化和硫代化(PCP)结构域之间的融合,设计并合成二肽形成的模板,已在体外合成了预测的二肽。人工 NRPS 合成二肽速率与天然模板相似,NRPS C 端硫酯酶(TE)结构域可以作为肽环化的催化剂。类似聚酮合酶组装中的环化酶产生红霉素和酮类抗生素以及其他的大环内酯[13]。

5.3.2.4　模块募集

生物合成途径的进化是一个有迷惑力的问题,因为途径终产物往往用于调节途径中早期的酶活性。有几种理论,包括退化(retrograde)假说、正向进化、突变酶的逐渐积累和分子复原等已经在某种程度上指出生物合成途径如何进化。Jensen 假说指出新酶功能是由催化相似反应的酶的复原产生的[15],支持这个假说的是扁桃酸外消旋酶和黏康酸内酯化酶的结构和活性的比较。这两种酶在结构上是同源的,并且两者都催化由羧酸提取 α-质子。另一种酶,半乳糖酸脱氢酶与扁桃酸外消旋酶有明显顺序同源性,并且也催化由羧酸提取 α-质子所启动的反应。这就是说,与底物专一性明显趋异的一些同源酶共存于一个细胞之中。

了解自然界如何募集原始酶功能并使它们能够胜任其他任务,可以用于设计

具有新性质的酶。在自然界,新基因可以通过外显子改组产生。这表明特异功能可以复原和同其他基因组合产生新酶。*E. coli* 从头生物合成嘌呤的第三步是由甘氨酰胺核苷酸(GAR)转甲酰酶催化的。这个 23kDa 酶催化甲酰基由 N^{10}-甲酰基四氢叶酸转移到 GAR 给出甲酰基-GAR 和四氢叶酸。*E. coli* 第二个能使 GAR 甲酰化的酶是 *purT* 基因的产物,其与 purN GAR 转甲酰酶有很少的相似性,类似的酶已在 *Bacillus subtilis* 中发现,但至今没有鉴定出真核生物同源物。purT GAR 转甲酰酶用 ATP 驱动正向 GAR 甲酰化(用甲酸作为甲酰基供体)。甲酸被认为至少部分由另一个惟一的原核生物 N^{10}-甲酰四氢叶酸水解酶产生的,该酶由 *purU* 基因编码,它水解 N^{10}-甲酰四氢叶酸为甲酸和四氢叶酸。在 DNA 水平上,在 purN GAR 转甲酰酶的 N^{10}-甲酰四氢叶酸结合结构域和 purU 的 C 端区之间有大约 60% 同源性。这两个酶提供体外分子募集(molecular recruitment)研究的模型系统。总之,通过 purN 的 GAR 结合结构域和 purU 区的募集产生 GAR 转甲酰酶活性,催化 N^{10}-甲酰基四氢叶酸的结合和水解。

应用分子募集和外显子改组的思想产生杂合 GAR 转甲酰酶,其活性是天然 purN GAR 转甲酰酶的 100 ~ 1000 倍[16]。该工作指出 purN GAR 转甲酰酶可以通过募集 GAR 结合结构域使其能够结合和水解 N^{10}-甲酰基四氢叶酸。酶设计的第一阶段是催化结构域同定位受体底物的结构域融合产生催化剂;第二阶段可以把催化的残基工程到具有结合活性的蛋白质中产生酶;第三个阶段,把结合和催化的两种残基引入结构框架中,例如 β/α 桶蛋白。

5.3.3　杂合酶[11]

5.3.3.1　杂合酶的概念

近年来,有关杂合酶(hybrid enzyme)的研究日益受到重视,已成为分子酶学工程研究的重要领域。虽然许多文献中都有"杂合酶"的提法,但直到目前还没有一个统一而准确的定义,如果给杂合酶以定义的话,我们认为应该从结构与功能两个方面来考虑。这样,杂合酶可以定义为不同酶分子中的结构单元(二级结构,三级结构,功能域等),或整个酶分子进行组合或交换,从而产生具有新酶活性的杂合体。

Hopfer 将凝血因子的 N 末端亚结构域与胰蛋白酶 C 末端的亚结构域重组在一起,创造出一个有新活性的酶,这个酶可以使水解的底物范围变宽,并且表现出一些不同于亲本的新性质。

另一个构建杂合酶的例子是对酰基载体蛋白脱氢酶(acyl carrier protein hydrolase, ACP 脱氢酶)的定向进化研究。ACP 脱氢酶可以在一定链长度的脂肪酸的特定位置引入双键;植物油中单一不饱和脂肪酸的高含量,主要是由这些不同底物专

一性的 ACP 脱氢酶造成的,为了理解 ACP 脱氢酶底物专一性不同的结构基础,人们构建了一系列杂合酶,由 Δ^6-16:0-ACP 脱氢酶和 Δ^9-18:0-ACP 脱氢酶相互交换一段顺序构成。经过筛选产生的杂合酶,表现出底物链长专一性或引入双链位置特异性的改变,其中一个杂合酶比活性比野生型 Δ^6-16:0-ACP 脱氢酶高:对 16:0-ACP 比活性增加 2 倍,对 18:0-ACP 比活性增加 15 倍多。

在许多情况下,同源区域间残基或结构交换而获得的杂合酶性质介于双亲酶之间,活性降低。相似性越低,杂合酶活性越低甚至没有活性,然而利用随机突变的方法可以恢复杂合酶的活性。

5.3.3.2 构建杂合酶的方法

有许多方法可以产生杂合酶,如 DNA 改组,不同分子间功能域交换,甚至两个酶分子的融合等,或者把来自不同酶分子的(亚)结构域进行重组成为一个新的单一结构域,或者把来自不同酶的本身没有活性的模块重组起来,同时在一个进化体系中筛选,就有可能获得比亲本催化效率更高,或者衍生出新功能的杂合酶。现将构建杂合酶的方法介绍如下:

(1) 基因随机缺失法

我们利用基因随机缺失法,构建了天冬氨酸酶和 α-天冬氨酰二肽酶的杂合酶,其基本方案是利用核酸外切酶 Ⅲ 将天冬氨酸酶基因的 C 端随机缺失一部分,将 α-天冬氨酰二肽酶基因的 N 端随机缺失一部分,并构建缺失基因文库。然后,将这两个库连接成为杂合基因库,在特定的选择压力下,从受体菌中筛选杂合酶,得到了具有 7%天冬氨酸酶活力和 26% α-天冬氨酰二肽酶活力的杂合酶。这两个酶基因序列之间基本上没有同源性,而且在蛋白质结构上也相差较大。天冬氨酸酶是一个四亚基酶,而 α-天冬氨酰二肽酶是单链酶。这个杂合酶的构建表明基因随机缺失法是构建无同源序列多亚基酶和单链酶的杂合酶的有效方法,为构建杂合酶提供了一个新途径。例如我们可以将目前采用多酶处理的多步反应工艺过程,利用杂合酶将反应的步骤减少,甚至变为一步反应,这样将会大大降低工业成本。

(2) 同源基因改组

同源基因改组可以产生具有新性质的杂合酶,这是构建杂合酶的一个重要方法。经常采用同源扫描突变,所谓同源扫描突变(homologue-scanning mutagenesis)是几个同源酶基因应用 PCR 扩增,然后用内切核酸酶分裂,不同的分裂产物混合并一起退火,不同基因的同源区组合,用 Taq 聚合酶扩增,含有几个同源酶的基因片段组成新基因-杂合基因,进而克隆和表达,产生杂合酶。

(3) 非同源基因改组

很多酶很少或者没有 DNA 的同源性,可能有蛋白质结构的同源性。构建这样

蛋白质结构同源性的杂合酶可能是产生杂合酶另一重要方法。

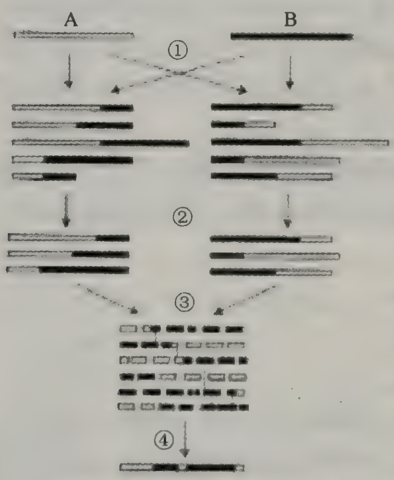


图 5.16　非同源改组

①构建(A-B)和(B-A) ITCHY 库(见 3.13);②用外切限制酶或 PCR 引物获得与出发基因同样大小的基因;③两个 ITCHY 库的基因混合,用 DNase I 片段化,DNA 改组;④通过模板转换组装随机片段,筛选

非同源基因的改组,它是由两个 ITCHY (incremental truncation for the creation of hybrid enzymes)库组成的:一个在 N 端(A-B)带有基因 A,一个在 N 端(B-A)带有基因 B。然后将分离的 A-B 和 B-A 的各种 DNA 片段融合,它们与最初的基因大小几乎相等。为了改组需用 PCR 扩增,两个库(A-B 和 B-A,PCR 产物基本与原基因大小相同)被混合,用 DNase I 消化和 DNA 改组(图 5.16)该法较传统 DNA 改组产生多样性更多的库。

(4) 增长缩短法

增长缩短是通过缓慢的、定向控制 DNA 的消化。这样可以产生所有可能单一碱基对缺失的 DNA 片段库。

Luts 等报道了构建杂合酶的“增长缩短法”(incremental truncation for the creation of hybrid enzyme, ITCHY)。它是根据在两个目标基因片段之间产生单一碱基对缺失库的思想,继之这两个片段库随机融合(图 5.17)。该法作为模型系统已成功地用于 *E. coli* 甘氨酰胺核苷酸转移酶(purN)和人(hGART)杂合体。作者应用 ITCHY 作为起点,导入 α-硫代核苷酸类似物(αS-dNTPs)以产生增长缩短库(THIO-ITCHY)。这个方法是根据目标基因 PCR 扩增时 αS-dNTPs 随机低频植入,然后外切酶处理,结果形成在随机位点水解终止,并一步产生缩短库。

(5) 应用蛋白质内含子构建杂合酶

蛋白质内含子(intein)是蛋白质中的一段多肽链,靠自我剪切的方式从前体蛋白质分离出来,同时两端的蛋白质外显子(extein)以肽键的方式相连(图 5.18),序列和结构分析比较表明蛋白质内含子的 N 端大约有 100 个氨基酸残基和 C 端大约有 50 个氨基酸残基对自我剪接至关重要,其间由一个连接区序列分开。

根据 http://www. neb. com/inteins html 数据库,蛋白质内含子有 8 个模体。其中,A、B、F、G 属于自我剪接模体[图 5.18(b)],而 C、D、E、H 属于自我剪切模体。模体中没有绝对保守的氨基酸残基。模体 A 中 Cys/Ser/Thr(图 5.18 中 C/S/T),模体 B 中 Thr * * Ser,模体 G 中 His-Asn 相对保守。

蛋白质内含子自我剪切过程示如图 5.19。总反应过程可以看作亲核反应。

蛋白质自我剪接是蛋白质在后翻译水平上的加工过程,即蛋白质内含子从前体

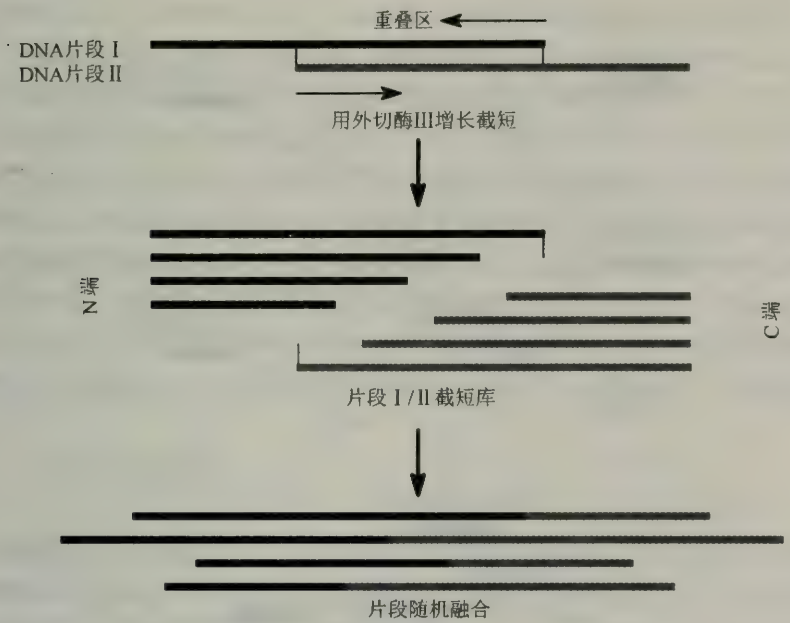


图 5.17 在两个 DNA 片段之间产生增长缩短库一般图解

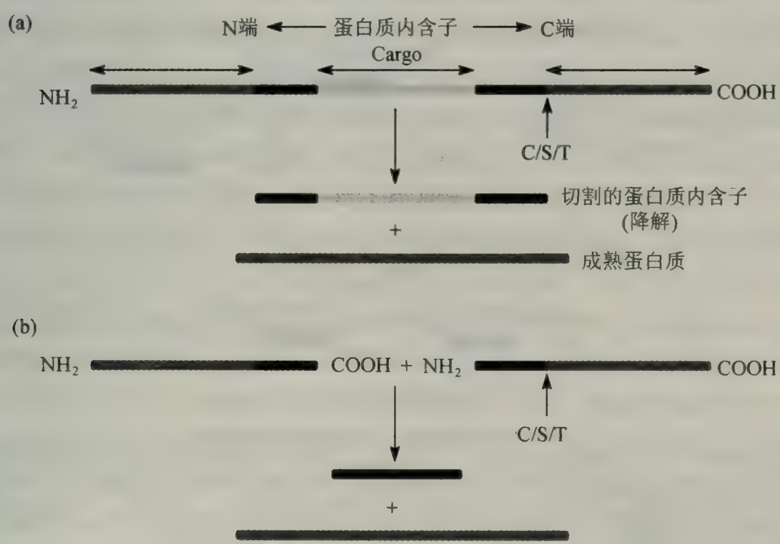


图 5.18 蛋白质内含子的结构和成熟

(a)原型结构，包括内肽酶顺序之间的 Cargo；(b)剪接蛋白质内含子，编码两个分离的转录单位，蛋白质产物连接形成功能蛋白质内含子，然后切除

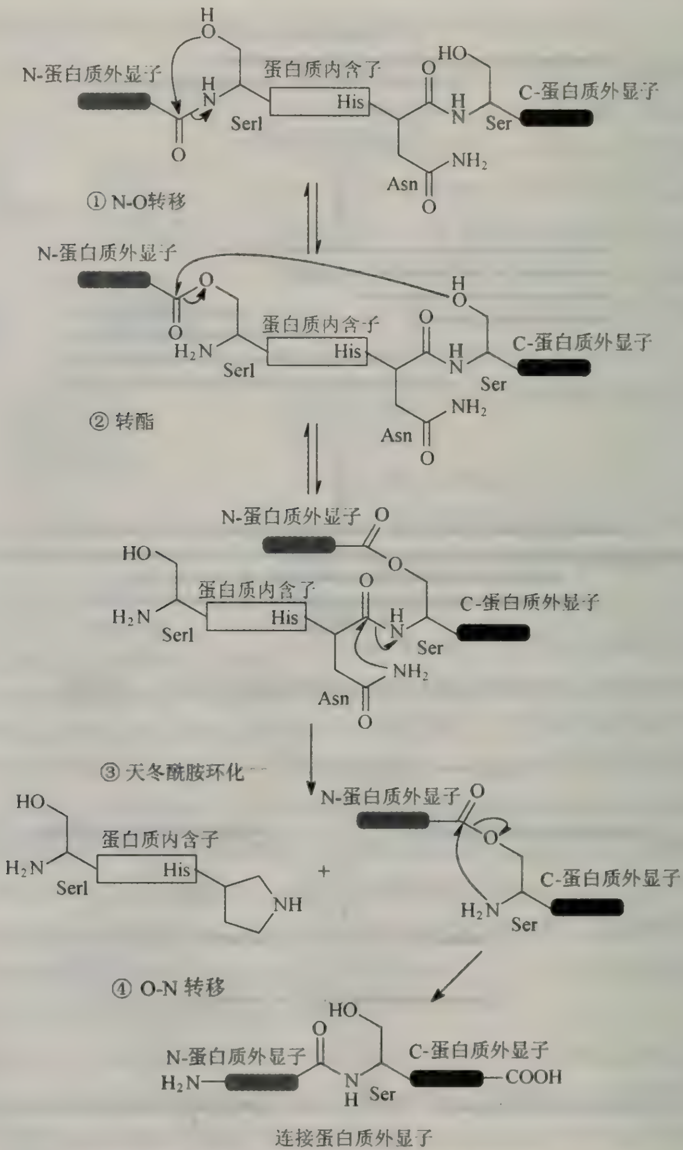


图 5.19 蛋白质内含子自我剪切过程

① 由于蛋白质内含子 N 端的酰胺键，在模体 B 中的 His Thr 作用下形成不稳定的顺式构象，而与蛋白质内含子的 N 端 Ser 侧链上羟基发生反应导致蛋白质内含子的 N 端外显子转移到 Ser 的侧链上。已发现一些蛋白质内含子 N 端是 Ala 残基，既然 Ala 不含羟基或硫基的侧链，就不能像 Ser/Thr/Cys 一样能够完成酰基转移，但如果 N 端连接肽键不稳定，而下游外显子第一个氨基酸残基在适当的位置可以攻击这个肽键，则可以直接将上游外显子转移到下游外显子的第一个氨基酸残基侧链上。② 下游外显子的 Ser/Thr/Cys 攻击上游外显子的酯/硫酯键，导致上游外显子从蛋白质内含子的 N 端分离，而转移到下游外显子的 Ser/Thr/Cys 的侧链上，从而形成分支蛋白。③ C 端保守残基 Asn 环化形成琥珀酰亚胺环，导致 C 端剪切位点的断裂，琥珀酰亚胺能够水解形成 Asn 或者 isoasparagine。有些蛋白质内含子 C 端是 Gln，虽然这些蛋白质内含子还没有发现它们的自我剪切活性，但如 Asn 一样，Gln 也能发生环化反应从而可以取代 Asn 而完成这一功能。④ O-N 或 S-N 酰基重组，在上下游外显子之间形成天然的肽键，这一过程是自发进行的

蛋白质中剪下来,其两侧的蛋白质外显子的 C 端区和 N 端区连接,产生两个新的蛋白质。这就是说由一个基因可读框产生两个以上蛋白质,而且其中一个蛋白质的基因编码和最初的可读框不成线性关系，从而开辟了构建杂合酶的一个新方法。

在新酶的开发和现有酶的改进利用方面,杂合酶技术显示了它巨大的应用前景。仅 1998 年就有 14 个采用杂合酶技术改进的酶获得了美国专利。我们有理由相信:不久的将来，杂合酶技术将会在分子酶学工程研究和开发方面展现自身的价值。

杂合酶技术是将 DNA 水平上的突变筛选与蛋白质水平上的酶学研究相结合的一门综合技术。它将传统酶活性筛选法同简便的 DNA 重组技术有机地结合起来。这一技术的引入,使分子酶学工程的研究摒弃了繁琐的蛋白质序列研究和繁重的菌种选育这一传统做法，为加快构建新酶和改进生物工艺过程开创了一条新路。

5.3.4　定向共进化

海藻糖是一种非还原性二糖,通常作为应激代谢物而广泛存在于生物体内。近年来因其对生物大分子的非特异性的保护作用而备受人们关注。在微生物体内,海藻糖的体内合成主要通过三种途径,其中大多数微生物广泛采用的是通过一个双酶体系——海藻糖-6-磷酸合成酶/海藻糖-6-磷酸水解酶酶系完成。海藻糖-6-磷酸合成酶将高能糖化合物 UDP-葡萄糖的葡萄基转移到葡萄糖-6-磷酸上合成海藻糖-6-磷酸,再由海藻糖-6-磷酸水解酶水解掉磷酸生成海藻糖。在大肠杆菌中,该酶系由位于基因组 42min 的操纵子 *otsBA* 编码。该操纵子全长 2.1kb,基因 *otsB* 位于上游,编码海藻-6-磷酸水解酶,而基因 *otsA* 位于 *otsB* 下游,编码海藻糖-6-磷酸合成酶,二者之间有 23nt 的重叠。鉴于该酶系在工程学领域的意义及其基因一级结构的特殊性,我们对该操纵子进行了定向进化。将两个基因作为一个整体进行进化和筛选,最终找到两个酶的高产海藻糖最佳配合条件。我们将这种定向进化称之为定向共进化[17,18]。这样就可以将两个基因放在一起考虑,而不必考虑当一个酶达到最佳状态时是否会对另一个酶产生不利的影响。首先克隆 *E. coli* 操纵子 otsBA,并用易错 PCR 对该操纵子随机突变;然后,PCR 产物混合物直接通过新的 DNA 改组方法——shuffling PCR 进行重组(图 5.20)。Shuffling PCR 方法的主要思想是缩短延伸反应时间,并使截断的 PCR 产物在下一循环中作为引物,随机结合模板库中的模板。与 Stemmer 的 shuffling 方法不同,shuffling PCR 方法中改组与扩增都是在一步 PCR 操作中完成。这样使点突变的重排更加快速和高效。我们又通过紫外诱变构建了一株海藻糖缺陷型菌株,*E. coli* DEF3(tre^-),并将其作为重组质粒的宿主菌,以消除菌体对海藻糖利用所造成的影响。

对大约 4000 个菌落进行了筛选,共有 15 株表现出高的海藻糖含量。其中 *E. coli* TS7 是海藻糖产量超过野生型操纵子 otsBA 3.7 倍的突变体,该突变体海藻糖

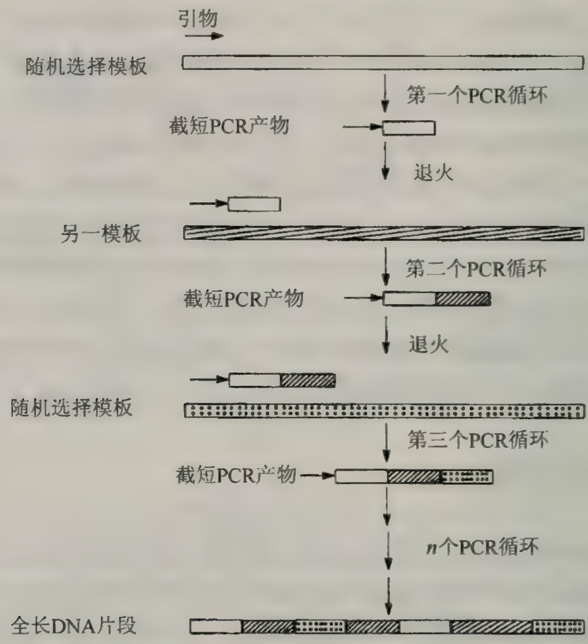


图 5.20　shuffling PCR 操作

产量较 *E. coli* DH5α 高 12.3 倍。蛋白质电泳分析表明，突变体与野生型操纵子在 *E. coli* 中表达量接近，证明海藻糖含量的提高主要是酶活性提高的结果。对 *E. coli* 操纵子 otsBA 的进化结果不仅为工业化生产海藻糖提供了合适的基因工程菌，同时，也表明所设计的定向共进化方法的成功。

5.3.5　适应性进化

用定位诱变和定向进化所得到的工程酶都是自然界不存在的进化酶，其对环境的适应性是人们关注的问题。了解适应机制，或许可以作为工程酶的一种新方法。

在大多数情况下，酶适应极端环境不需要完全新的分子结构。许多极端嗜热酶（extremophilic enzyme）和嗜常温酶（mesophilic enzyme）适应环境是相似的，在不同的温度下有相近的 3D 结构（图 5.21）。顺序比较表明共同祖先的酶可以适应相似的环境。进化酶的研究也表明它可以迅速适应新的环境和较宽的温度范围。嗜常温酶和嗜冷酶（psychrophile enzyme）的高热稳定突变体在稳定性上相当于许多天然嗜热酶，并且抗蛋白酶水解和化学变性。定向进化也得到了嗜常温和嗜热酶的冷活性突变体。

类枯草杆菌蛋白酶的蛋白质水解酶（subtilisin-like protease）适应环境广，因此

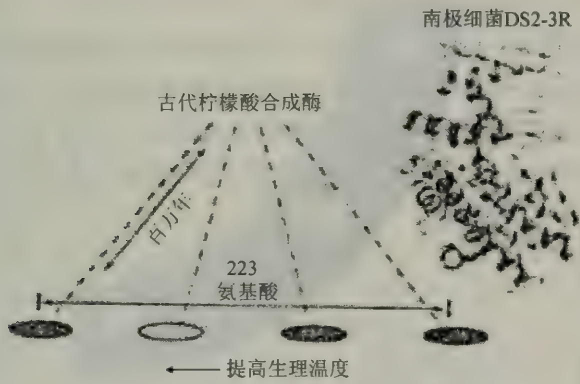


图 5.21　分支进化的酶对不同环境的适应性

这里以柠檬合成酶的亚基为例，它的嗜热酶和嗜冷酶有 223 个氨基酸是不同的。不同机体的现代柠檬酸合成酶能在较宽的温度范围下生存，但它们都起源于一个结构相同的祖先。尽管进化的距离大，但它们的三维结构仍然非常相似。一些进化或是中性的或是接近中性的酶，序列的改变可能仅仅导致酶功能特性的很小变化。这样，很难用比较方法来研究适应机制

这族酶成为研究适应性的典型材料。*Antaretic bacillus* TA 41 的枯草杆菌蛋白酶 S41 在低温下具有高活性。研究发现 S41 有两个突出的特征：①有伸出的突环(loop)，而普通的嗜常温枯草杆菌蛋白酶，例如 BPN′未发现有此突环；②表面 ASP 残基含量高(22 个)，这可能表明嗜冷酶的特殊性质。由数据库中发现球形芽孢杆菌(*Bacillus sphaericus*)的嗜常温枯草杆菌蛋白酶 SSII 与 S41 非常相似，也有伸出的突环和 19 个 ASP 残基[19]，然而，SSII 有嗜常温酶的稳定性。由此可见，既不是伸出的突环也不是高 ASP 残基含量可以说明嗜冷 S41 的特殊行为。

适应机制目前尚不清楚，有物理-化学之说，也有生物学行为之解释，后者更为复杂。在进化酶的稳定性和活性观察中，似乎适应性是一种物理-化学机制[20]。例如，酶的热稳定性和催化活性之间的表观交替换位(图 5.22)，即酶在适应进化过程中，不断调整相互作用强度，在生理相关的温度下使刚性(对稳定性)和柔性(对活性)之间的平衡最佳化。这种对同源酶温度适应性解释似乎是合理的。理论上讲，所有酶都是进化的产物，应当遵循物理-化学规律。这样，我们似乎可以把酶在确定环境下的适应性称之为“定境适应”(environment-directed adaptation)。

冷适应酶的研究是有限的，但也获得了一些结果。大多数适应酶通过增加 K_{cat} 和 K_m 来增加催化效率。K_{cat} 增加是为了降低活化能，增加 K_m 可以说明有更柔性的构象。然而大多数冷活性酶伴随热稳定性降低。根据冷适应酶的晶体结构、同源模型和氨基酸顺序比较没有提示存在与冷适应相关的共同决定子。每个

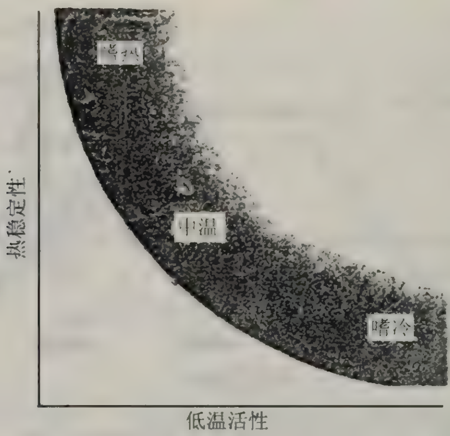


图 5.22　同源酶对各种环境的适应性

图为低温下同源酶的催化活性与热稳定性之间的表观交替换位。嗜冷酶低温下催化活性是高的，而嗜热酶通常是低的；热稳定性的关系正好相反：低温下，嗜热酶的热稳定性高，而嗜冷酶的热稳定性低

酶和酶类似乎有自己的适应策略。冷活性酶在很大程度上通过最佳化活性部位和活性部位周围静电相互作用来增加它们的催化活性。

天然酶在有机溶剂中，即使有时能保持天然构象但也极易失活。因此根据应用环境对酶分子定向进化就十分必要。例如进化的枯草杆菌蛋白酶在 60%DMF，提高了酶的活力和稳定性。进化的对硝基苄基酯酶在 30%DMF 中酶活力提高 100 倍[21]。有时候需要将常温生物中的酶在低温中应用，这就需要提高常温酶在低温条件下的活力。Bullow 等[22]以 NaCl 作为筛选压力，同时定向进化冬鲽抗冷冻蛋白 HPLC-8 组分（与蛋白 A 中的两个结构域融合）对高盐浓度和对低温的耐受力，得到的 4 个耐高盐浓度的突变体在低温下活力也有不同程度的提高，说明这两种性质有很大的相关性。Kano[23]首先利用随机突变来提高中温酶在较低温度下的活力，他在 10℃筛选枯草菌蛋白酶 BPN′的活性，使得 K_{cat}值达到了常温野生型酶的级别，但是它的 K_m 值却急剧下降，10℃水解速率增加了 10%，1℃水解速率增加了 30%。在更高的温度下稳定性却没有改变。

天然酶在生物体内不会接触到人工添加的去离子螯合剂，而洗涤酶必须在螯合剂存在的环境中保持活性。因此有必要改变蛋白酶依赖金属离子来保持其活性或稳定性的性质。Strausberg 等[24]进化了一株在缺少 Ca^{2+} 仍稳定的枯草杆菌蛋白酶。他们随机突变了 10 个氨基酸残基，删除了结合 Ca^{2+} 的突环。在无 Ca^{2+} 的溶液中进化酶半衰期比野生型 BPN′长 12.5 倍。

α-淀粉酶（α-amylase），特别是微生物 α-淀粉酶，已广泛用于工业过程，例如淀

粉液化和成浆过程，以及最近用于去污。由于 α-淀粉酶在碱型 pH 需要具有高活性和特异性（这是关键参数），特别是用于去污过程，Bessler 等[25]开展了细菌 α-淀粉酶定向进化研究。他们对野生型淀粉液化芽孢杆菌（BBA）及突变体 BAAS201N 和 BAAN297D 进行了易错 PCR 和基因改组。为了筛选突变体，他们开发出一种新的，可靠的适合于高通量筛选的方法（图 5.23、图 5.24）。突变体 BAA42 在 pH7 有

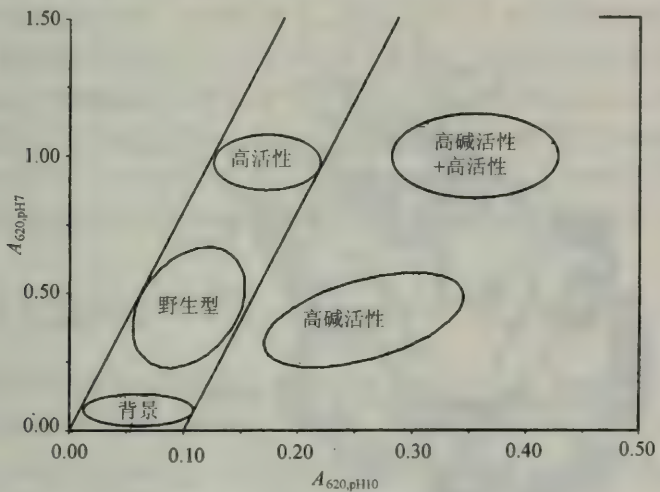


图 5.23　克隆双向筛选系统

α-淀粉酶活性是在 pH10 和 pH7 测定的

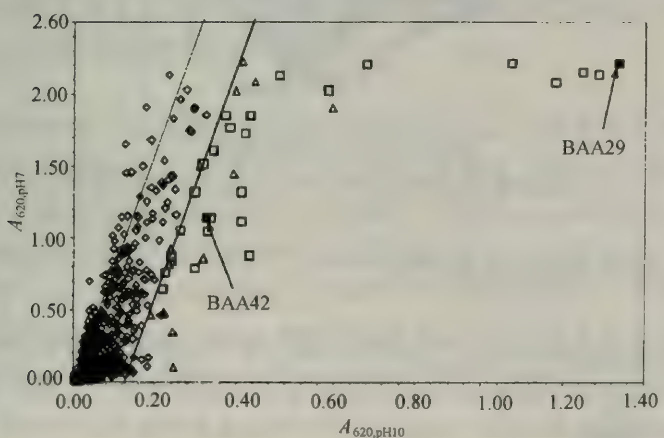


图 5.24　DNA 改组库的筛选

◇——DNA 改组库；□——DNA 改组库；△——易错 PCR 库

BBA29 和 BAA42 是两个突变体

最佳活性,与野生型比,pH 移动 1 个单位。BAA42 在较广 pH 范围有活性,在 pH10 活性提高 5 倍,周质提取物的活性和比活性分别提高 4 和 1.5 倍。另一突变体 BAA29 具有类野生型 pH,而周质提取物活性提高 40 倍,比活性提高 9 倍。两个突变体的氨基酸同其他的同源微生物 α-淀粉酶的氨基酸顺序比较揭示了高度保守的残基发生了突变,即 W194R、S197P 和 A230V(图 5.25)。

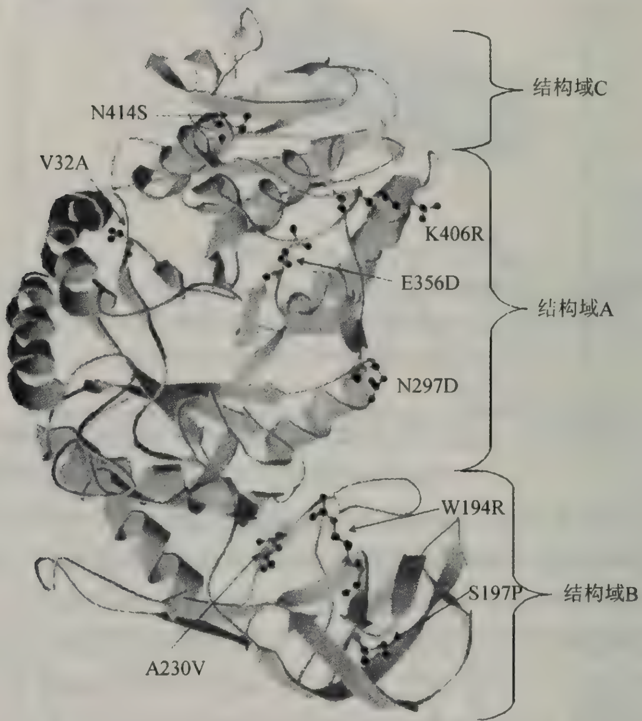


图 5.25 BAA 结构的突变

该结构是根据 BLA 同源模型化应用 Swiss Model Server 产生的

5.3.6 稳定性和活性的进化[26]

用定向进化改进酶的稳定性和活性在酶工程中占有重要地位。对应用来讲,酶的稳定性和活性非常重要,因此也是酶的定向进化非常重要内容。定向进化已成功地提高嗜常温或嗜冷酶的热稳定性或在最适温度时提高其活性或提高嗜热酶的低温活性而不影响其热稳定性。例如,将嗜常温枯草杆菌蛋白酶 E 进化为相当于其嗜热同源物的 thermitase(43%氨基酸同源性),最好的突变体仅有 8 个氨基酸改变,最适温度提高 17℃,在 65℃半衰期提高 200 倍并保持活性。*Therms thermophilus*

核糖核酸酶热稳定突变体在 40℃时 K_{cat}/K_m 提高 40 倍。超级热稳定 β-糖苷酶，嗜热吲哚甘油磷酸合成酶可以进化为低温活性突变体。然而低温活性往往伴随突变体稳定性明显降低，这就引发出一个问题，是否低温活性和高温稳定性不相容。研究表明这两种性质可以互相单独进化。酶两种性质的相关性研究认为可能是筛选的结果（仅仅选择一种性质）。嗜常温枯草杆菌蛋白酶（SSII）的随机突变产生的第一代突变体同野生型比较，在 10℃时 K_{cat} 提高 2 倍，在 70℃时半衰期增加 1.4 倍；热稳定性更高的突变体在 10℃时 K_m 和 K_{cat} 值表明，低温催化转换率所要求的构象柔性仅局限于活性部位，底物结合水平降低了，而总的稳定性或柔性不降低。关于在稳定性和活性中酶柔性的内涵用进化的热稳定的对硝基苄基酯酶突变体已进行了详细的研究[27]。在嗜冷枯草菌蛋白酶 S41 体外进化过程中，应用提高热稳定性和低温活性选择压力筛选出催化活性和热稳定性都提高的突变体。

5.3.7　底物特异性的进化

两种酶如果有紧密相关的催化机制或结构，那么，由一种酶进化为另一种酶，即产生底物特异性的新酶的概率就可能高。例如脂肪酶和酰胺酶，两种酶的底物通过 Ser-His-Asp 三元组转变。如果酰胺酶底物被脂肪酶接受或反则亦然，产生底物特异性的新酶的机会就多。

底物特异性定向进化可分为以下 3 组（图 5.26）：①不能转化底物。原则上在研究的条件下酶可接受底物，但由于立体障碍，电子效应等不能转化底物。②可转化底物。底物转化通常是由紧密相关的其他酶催化。③目标酶活性低。目标酶可以转化底物，但同天然底物比活性低。

组 1

Fersht 等由引哚-3-甘油磷酸合成酶（IGPS）的 $(\beta/\alpha)_8$ 桶骨架进化磷酸核糖酰基邻氨基苯甲酸异构酶（PRAI）活性[28]。它们成功地提供了酶进化“保守骨架”假说的典型范例。PRAI 的突环系统设计到 IGPS 上，用 DNA shuffling 和 StEP 定向进化。活性突变体应用在色氨酸缺乏情况下不生长的 *E. coli* 菌株进行琼脂平板鉴定，突变体与 PRAI 有 28% 的同源性，与 IGPS 有 90% 的同源性。基因改组把涉及组氨酸生物合成的 ProFARI[（N'-5-磷酸核糖酰基）亚胺甲基-5-氨基咪唑-4-羧酰胺核苷酸异构酶]转为在色氨酸生物合成中催化同样反应的具有 PRAI 活性的突变体（图 5.27）。

组 2

DNA 改组已成功地把 *E. coli* LacZ β-半乳糖苷酶转为岩藻糖苷酶和将环糊精糖苷转移为淀粉水解酶[25,29]。根据生色岩藻糖底物每轮筛选约 10 000 克隆，鉴定出进化的岩藻糖苷酶，与野生型比较，对岩藻糖底物 K_{cat}/K_m 提高 10 ~ 20 倍。与

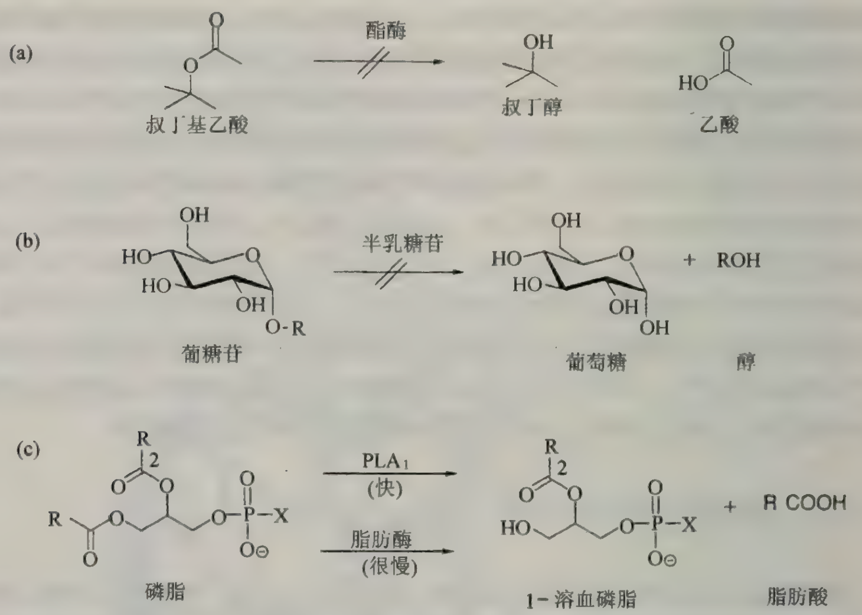


图 5.26 酶的新底物特异性可能的底物

(a) 各种酯酶不作用于叔醇的酯(组 1);(b) 各种半乳糖苷酶只接受半乳糖苷作为底物,而不接受葡糖苷(组 2);(c) 某些脂肪酶,例如 *Rhizopus* sp., *Bacillus* sp. 或 *Staphylococcus* sp. 脂肪酶,活性明显低于磷脂酶 A_1(PLA_1)(组 3)

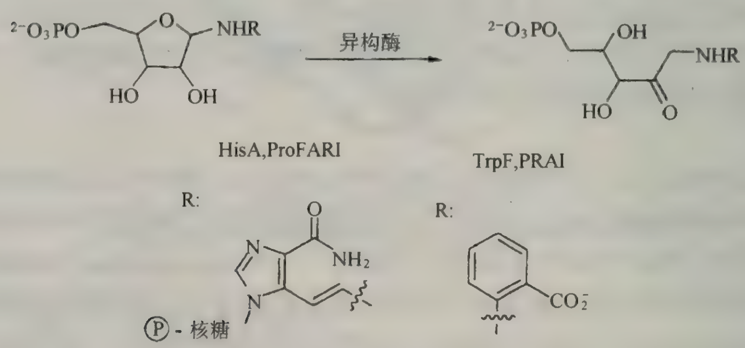


图 5.27 ProFARI(涉及组氨酸生物合成)和 PRAI(涉及色氨酸生物合成)的底物特异性

天然 β-半乳糖苷酶比较,最终进化酶底物特异性(邻-硝基苯基岩藻吡喃糖苷:邻-硝基苯基半乳糖吡喃糖苷)提高 1000 倍;同样对硝基苯酚衍生物底物特异性提高 300 倍。

定向进化将环糊精糖苷转移酶转为淀粉水解酶。α-淀粉酶族或糖苷水解酶是一大族淀粉加工酶。催化结构域的$(\beta/\alpha)_8$ 桶折叠,催化部位的残基和键分裂机制

是保守的,但产物和反应特异性是不同的。

环糊精糖苷转移酶(CGTase)是该族的一个成员,其形成环状 α-(1,4)糖苷链的

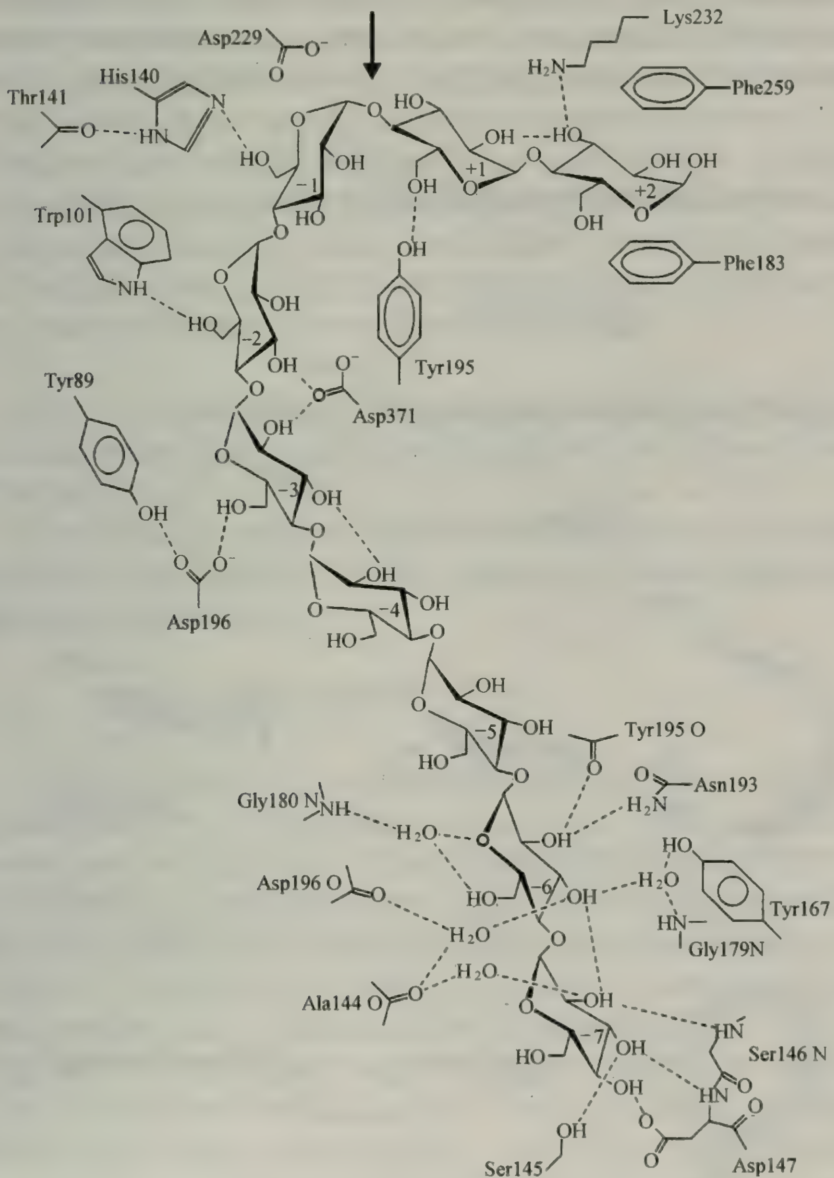


图 5.28　maltononaose 底物和环状芽孢杆菌 CGTase 结合底物裂缝之间相互作用图解

箭头指出切割亚位点 - 1 和 + 1 之间键;虚线表示氢键。Arg47 和 Asn94 不与未分裂的底物相互作用,但它们与反应中间体相互作用并产生 γ-环糊精。Phe183 和 Phe259 与糖环疏水堆积

寡糖(由 6 个、7 个或 8 个葡萄糖残基组成,分别为 α-环糊精、β-环糊精和 γ-环糊精)。它由 5 个结构域(A～E)组成;结构域 A 和 B 组成催化结构域,结构域 E 涉及结合生淀粉,而结构域 C 和 D 功能未知。在底物结合之后,穿过几个糖结合亚位点(图 5.28 中 –7～2),在亚位点 –1 和 +1 之间的 α-(1,4)-糖苷键分裂产生共价糖苷基-酶中间体,其结合在共体的亚位点(–1, –2, –3 等)(图 5.29)。下一步反

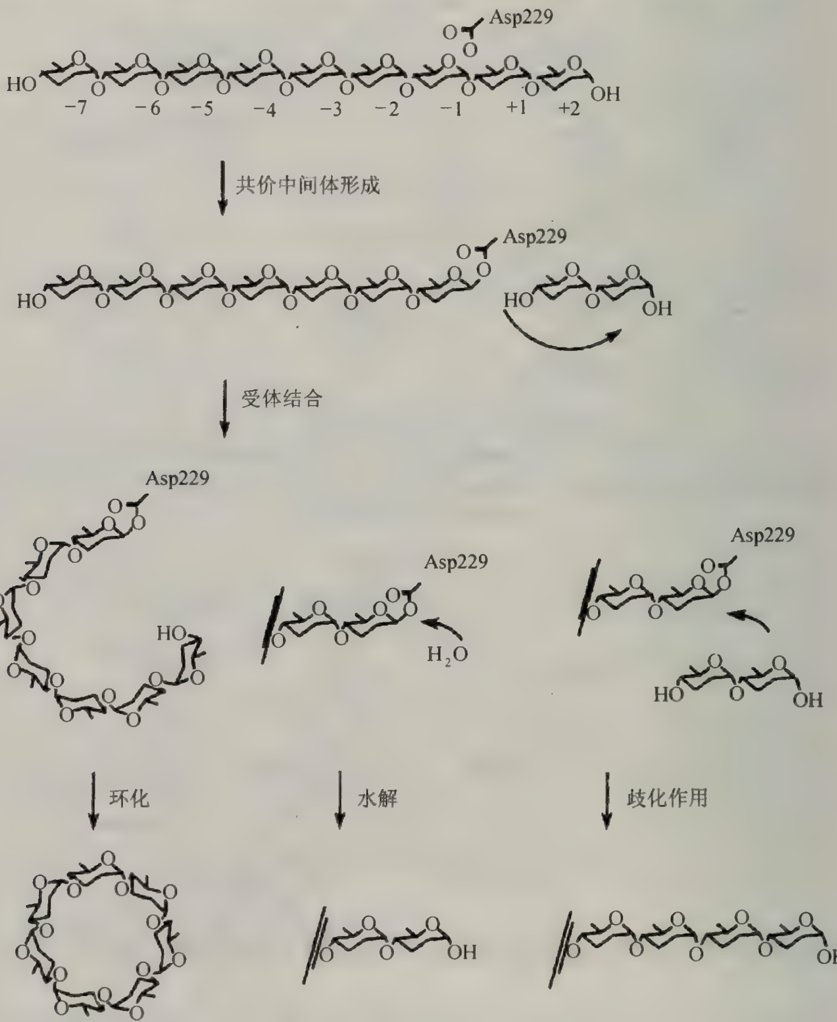


图 5.29　CGTase 催化反应的图解

在键分裂之后,共价结合反应形成酶-糖基中间体。反应的第二步,反应中间体被转移到受体分子。在环化反应中,共价连接的寡糖的末端 OH-4 基团作为受体,而水或第二个糖作为水解和不相称反应的受体

应,受体分子结合在受体亚位点 +1 并分裂糖基-酶之间的键。在环化反应中,共价结合糖的非还原端作为受体产生环糊精。CGTase 也可以用水或第二个糖分子作为受体,分别产生水解或不相称反应。水解和转糖基反应的相对效率是由用于反应的受体性质和受体亚位点的性质决定的。CGTase 对糖基受体明显优先,这是由于其转糖基化活性远高于水解活性。在受体亚位点 +2 突变可以把 CGTase 变为淀粉水解酶,然而,在其他亚位点突变对水解/转糖基化比率影响不大。为了解释转糖基化反应的特异性提出在受体亚位点 +1 结合糖而不是水。为了鉴定限制 CGTase 水解活性和转糖基化反应特异性的氨基酸残基,Dijkhuizen 等[30]对环状芽孢杆菌(*Bacillus circulans*)251(BC251)CGTase 随机突变并选择水解活性提高的突变体,明显提高 CGTase 的水解活性,并使其转为淀粉水解酶。在 3 轮易错 PCR 突变后,水解活性提高 90 倍,是至今报道最高活性的 CGTase。

组 3

野生型 *E. coli* 的醛缩酶几乎不接受非磷酸 D-2-酮基-3-脱氧葡萄酸(KDG)。几轮 epPCR 和改组鉴定出对 KDG 活性明显提高的突变体,最好的突变体催化效率提高 70 倍[31]。

上述的例子证明定向进化是改进酶底物特异性的有效工具;它可以解决传统筛选和培养方法存在的瓶颈问题;$(\beta/\alpha)_8$ 桶酶的进化也证实定向进化可以克服不同酶之间的功能和结构障碍。

5.3.8　对映体选择性的进化

手性有机化合物的立体选择合成具有重要的科学和工业意义。“手性市场”迅速扩展,仅 2000 年手性药物总销售额超过 1000 亿美元。目前,工业最常用的方法是经典对映体分离,但需要化学计算量的光学活性试剂和大量的有机溶剂。由于生态和经济原因,不对称酶催化预期会更有效,可以提供活性和高对映体选择的催化剂,催化非对称转化,或者外消旋化合物的动力学拆分或者前手性底物反应。两种催化剂——过渡金属催化剂和生物催化剂是有效的。非对称过渡金属催化需要有效配体的调节,这是一个困难的目标。酶、抗体酶和核酶已引起人们的关注。酶已用于某些手性中间体对映体选择的生产过程,例如脂肪酶催化环氧酯的动力学拆分已用于高血压治疗药物 Diltiazem 的生产。然而,这些催化剂对映体选择性可能很弱,与过渡金属催化中配体调节相似。定向进化已用于提高对映体选择性。该思想是从低对映体过量(ee)或选择性因子(E)值的野生型酶出发,以既定转化(A→B)为目标,建立突变体基因库,继之表达和筛选。

5.3.8.1 酶基因的突变、表达和酶分泌

建立定向进化方法的第一步是发展有效的表达系统，特别是超表达与酶分泌偶联。一个突出的例子是枯草杆菌蛋白酶。然而，许多目标酶不服从这样的系统，包括假单孢杆菌的各种脂肪酶。

编码细胞外酶的基因通常在标准宿主 *E. coli* 中超表达，但是蛋白质将在细菌胞质中形成失活的包涵体。为了解决这一问题，用革兰氏阴性菌绿脓杆菌（*Pseudomonas aeruginosa*）作为表达和分泌宿主。该菌株产生并分泌各种胞外酶，包括各种水解酶和假单孢杆菌的各种酶，它们折叠和分泌是高度特异的有规则的过程，至少已鉴定出 5 种不同的分泌途径，它们是由 3～20 种不同蛋白质组成的复杂细胞分泌机构。

构建的 *P. aeruginosa* 菌株 PABST7.1 带有强 T7 启动子，酶基因超表达。编码噬菌体 T7RNA 聚合酶的基因 *T7pol* 已克隆在启动子之后，该启动子可以被直接定位在 *T7pol* 下游 *LacI*^{q} 基因编码的 Lac 阻遏蛋白所抑制。这些基因稳定地整合到脂肪酶缺陷型 *P. aeruginosa* 菌株 PABS1 的染色体上（图 5.30），结果形成超表达菌株 *P. aeruginosa* PABST7.1。此外，构建的菌株 *P. aeruginosa* PAFR7.7 含有直接插

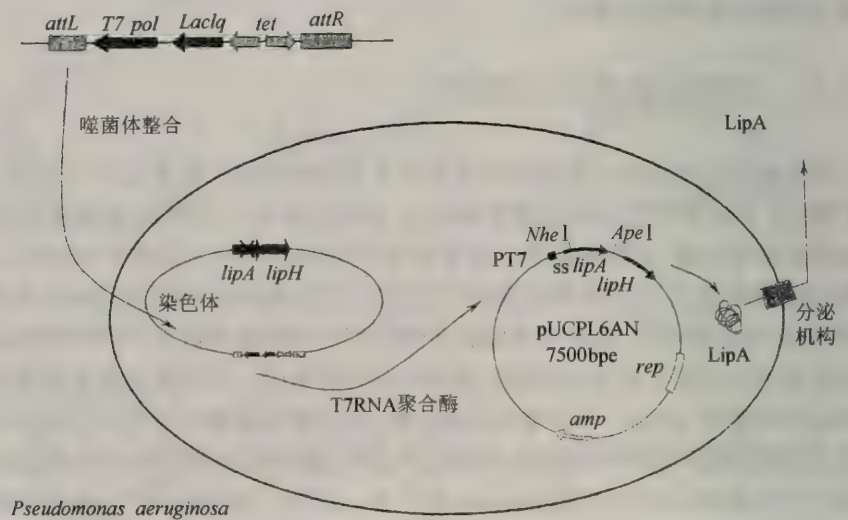


图 5.30 *Pseudomonas aeruginosa* 基因的超表达和酶分泌系统的构建

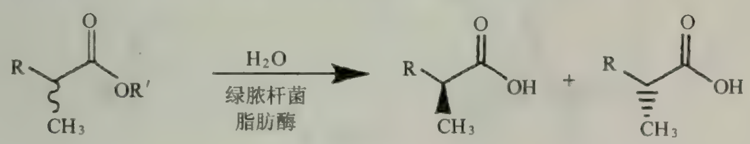
含有 *T7pol* 基因的噬菌粒被整合到脂肪酶缺陷菌株 *P. aeruginosa* PABS1 的染色体上。质粒 pUCPL6AN是由 pUCPKS 构建的。T7RNA 聚合酶的转录是通过加入异丙基-β-D-硫代半乳糖苷诱导，然后聚合酶指导脂肪酶操纵子的转录。脂肪酶正确地折叠并分泌到细菌培养上清液中，用于检验脂肪酶的活性

入脂肪酶结构基因 *LipA* 中的 *T7pol* 基因。两个菌株很适于突变体脂肪酶基因的超表达。应用该技术可以容易把 *T7pol* 基因定位到细菌染色体中,因此允许构建 *P. aeruginosa* 的超表达菌株以筛选定向进化的任何酶。

该系统的第二部分包括质粒 pUCPL6AN,其中脂肪酶操纵子 LipA/H 被克隆在 T7 启动子之后(图 5.30)。该质粒含有惟一 NheI 和 ApaI 位点,它们定位在脂肪酶结构基因 *lipA* 的 5′-和 3′。当构建物用于定向进化实验时,可以克隆任何突变脂肪酶的基因(这是特别重要的),并且突变引入既不影响信号顺序也不影响 3′调节区或特异的脂肪酶的折叠酶(foldase),因此,可以避免不需要的突变体表达。这个系统已成功地用于突变体脂肪酶大库的超表达和分泌。

5.3.8.2　脂肪酶突变体催化手性酯的动力学拆分[32~34]

脂肪酶是有机化学最常用的酶,催化羧酸酯水解或在有机溶剂中的逆反应。天然酶的对映体选择性是弱的,在约 50% 转化(*S*)-酸 10 时,有 5% ee(图 5.31)选择因子 E[反映(*S*)-和(*R*)-底物反应的相对速率]只是 1.1。



rac-9
R = *n*-C_8H_{17}

(*S*)-10　(*R*)-10

图 5.31　外消旋酯水解动力学拆分

当进行随机突变时,首先需要考虑拓展蛋白质的顺序空间、脂肪酶是由 285 个氨基酸组成,理论上有 20^{285} 种不同的突变体。按照结构改变的最小量,即每个酶分子只有一个氨基酸取代的话,根据 $N = 19^M \times 285!\ /[(285 - M)!\ \times M!]$,其中 M = 每个酶分子氨基酸取代数(这里 $M = 1$),突变体 N 库理论上有 5415 成员。然而,由于遗传编码的冗余度,通过 epPCR 产生库(含有全部的 5415 突变体)是不可能的。如果突变速率增加至这样一种程度,每个酶分子平均交换两个氨基酸(M = 2),那绿脓杆菌脂肪酶突变体数通过上述算法预计增加到 1.5 亿,即使应用目前有效的高通量筛选系统也很难检验。当 $M = 3$,突变体数是 520 亿。

最初的策略是根据相对低突变速率($M = 1$),逐步改进对映体选择性。当产生成千上万脂肪酶突变体(或任何其他的酶)时,这个问题是难以解决。然而这个问题从未出现,由于在琼脂板上所获得的细菌克隆,每个都来自单一细胞,这意味着每个细菌克隆只产生一个突变酶(有的可以产生一个以上)。然后细菌克隆通过克

隆采集器自动收集,并且在含有营养汤的微量脂酸冻点板上各孔中生长。在这种方式中,每个孔的上清只含一种突变体。

脂肪酶的定向进化过程见第 2 章中图 2.1。第一轮产生只有 1000 突变体的库,最好的突变体 ee = 31%,E = 2.1,第四轮,ee = 81%,E = 11.3(图 5.32)。当进行随机突变时,首先需要考虑开拓蛋白质的顺序空间,由一代突变体到下一代突变

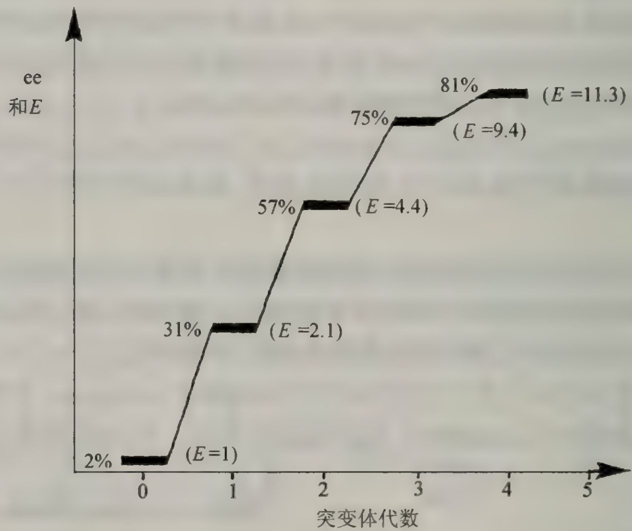


图 5.32 提高脂肪酶催化手性酯水解的 ee 和 E 值

体,在蛋白质顺序空间可能有许多不同“途径”。这样,如自然进化本身一样,有许多分枝的树(图 5.33),需要发现最短的可能路线爬上“ee 树”(或“E 树”)。用符号表示突变体的进化,其催化带有相反绝对构型产物的形成。虽然图中说明相当复杂,但也指出对于给定的反应,产生和发现高对映体选择的催化剂不是惟一的。这意味着应当尽可能的获得不同突变体的整个族,对给定的的反应都具有高对映体选择性。

一种策略不仅适用于脂肪酶催化的模型反应 9→10,而且也在其他对映体选择酶的进化上证明是有效的。这种策略是首先对每代最好突变体进行 DNA 顺序测定,以确定对映体选择性起作用的氨基酸取代的位置和性质。典型的资料如图 5.34 所示。脂肪酶中存在着改进对映体选择性的敏感位置(“热点”)。epPCR 和饱和突变的组合成为探索“热点”的主要方法。的确这个策略导致高(S)-选择突变体的产生(ee = 88% ~ 91%; E = 20 ~ 25),它们都是母体野生型脂肪酶派生的。为了进一步开拓蛋白质顺序空间,采用了 epPCR/DNA shuffling 和修饰组合多重盒式突变 CMCM(图 5.35)。后者是 DNA shuffling 的特殊类型,它可以产生突变基因库,

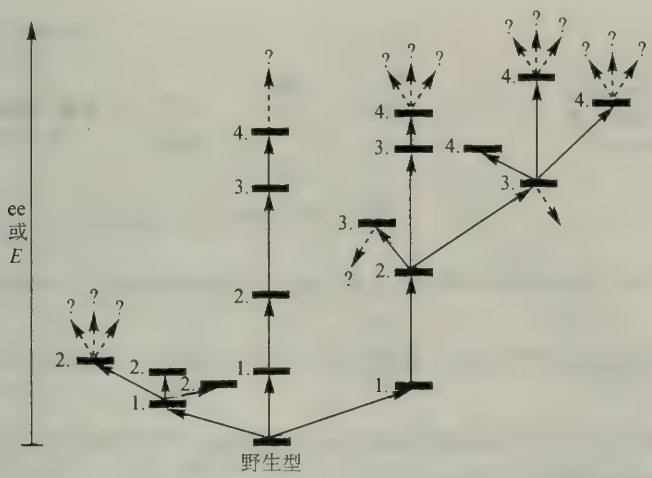


图 5.33　对映体选择性蛋白质顺序空间复杂度的“进化树”简化图解

数字表示各自世代突变和筛选得到的正突变酶。下指箭头表示

下线突变体(数字之外的少数正突变体)

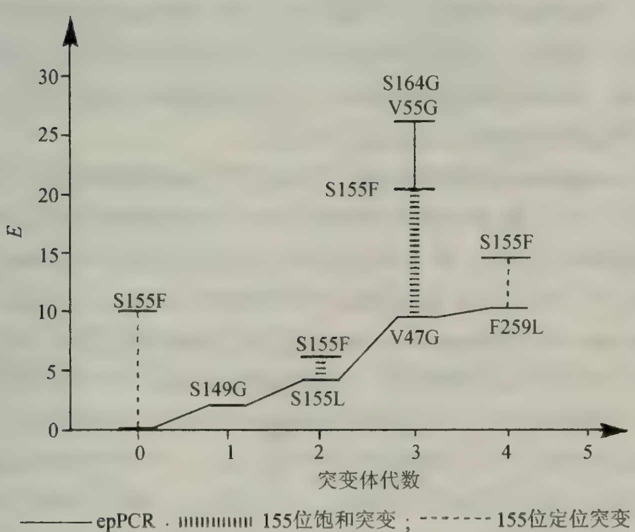


图 5.34　脂肪酶突变体催化酯 9 的对映体选择(氨基酸用单字母表示)

其中盒是由随机或确定顺序组成,并且野生型被随机地掺入。采用 40H4 + 38C8 两个突变体,在位置 155 和 162 饱和,继之表达和筛选,获得的突变体 1H8,$E>51$,ee > 95%,活性高且稳定。DNA shuffling 的几种形式相当成功。

另一个有趣的问题是关于对映体选择性逆向可能性。初步实验证明对映体选

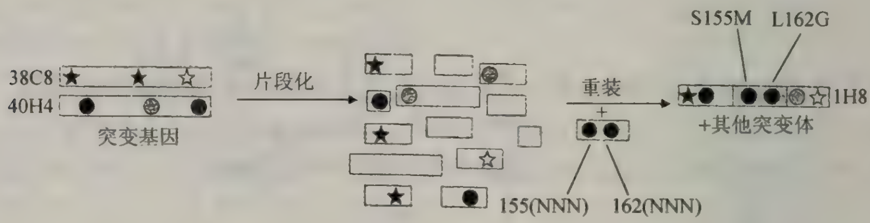


图 5.35 脂肪酶基因的修饰组合多重盒式突变图解(N 表示任一氨基酸)

择性逆向性是可能的(图 5.36)。由图可见,在同一库中,(*S*)-及(*R*)-选择突变体共存。

上述实验不仅代表对映体选择酶定向进化的第一个实例,而且也是对映体选择酶顺序空间开拓的最全面研究的例子。

P. aeruginosa 脂肪酶的 3D 结构以共价结合抑制剂复合物形式被解析(分辨率在 2.54 Å)。它有 α/β 水解酶折叠,由两侧 α 螺旋围绕的 6 股平行 β 折叠组成。催化三元组是由 Ser82、Asp229 和 His251 组成。含有 His251 的突环由八面体配位的钙离子所稳定。无或有底物决定这个酶取闭合或开放构象(酶结构是以开放构象解析的)。在这个构象中,可以鉴别出氧阴离子穴的位置以及可以容纳脂肪酸链三个底物结合口袋。保持底物酰基基团的主要力是范德华力相互作用。酰基口袋的大小和相互作用是酶对映体优选的主要决定子。

野生型 *P. aeruginosa* 脂肪酶的 3D 结构可以定位突变酶的取代位置和合理说明突变效应(图 5.37)。最佳的脂肪酶突变体(94G12)对底物(9)(*S*)对映体的选择性是由五个氨基酸取代表征,即 V47G、V55G、S149G、S155F 和 S164G,其中 4 个取代引入甘氨酸,它可以增加脂肪酶的整体构象柔性。柔性是决定野生型 *Geotrichun candidum* 和 *Rhizomucor miehei* 真菌脂肪酶底物选择性的重要参数。然而,这些酶的核心似乎明显是刚性的,而柔性结构域局限于少数表面的突环(此结果是分子动力学模拟得出的)。这样的骨架包含结构保守和刚性的 α/β 结构域以及少数暴露表面的突环。令人惊奇的是突变体 94G12 中 5 个取代的氨基酸没有一个在活性中心部位与结合底物直接接触,这样排除直接空间影响对映体选择性。各种取代直接定位或紧邻突环(突环涉及酶由闭合到开放构象的过渡)。在许多其他的对映体选择的突变体中也发现了相似的模型。空间上由活性部位去除氨基酸取代的遥控效应将影响酶的活性和稳定性,这是第一个观察到影响立体选择性的遥控氨基酸取代。

氨基酸 Val47 和 Val55 直接或间接涉及氢键网络,它把 His14 固定在螺旋 α2 和股 β4 和螺旋 α2 之间的突环上。His14 本身是含有涉及氧阴离子穴形成 Met16 突环的一部分(一种过渡态形成的结构前提)。因此,取代 V47G 和 V55G 可以预期增加

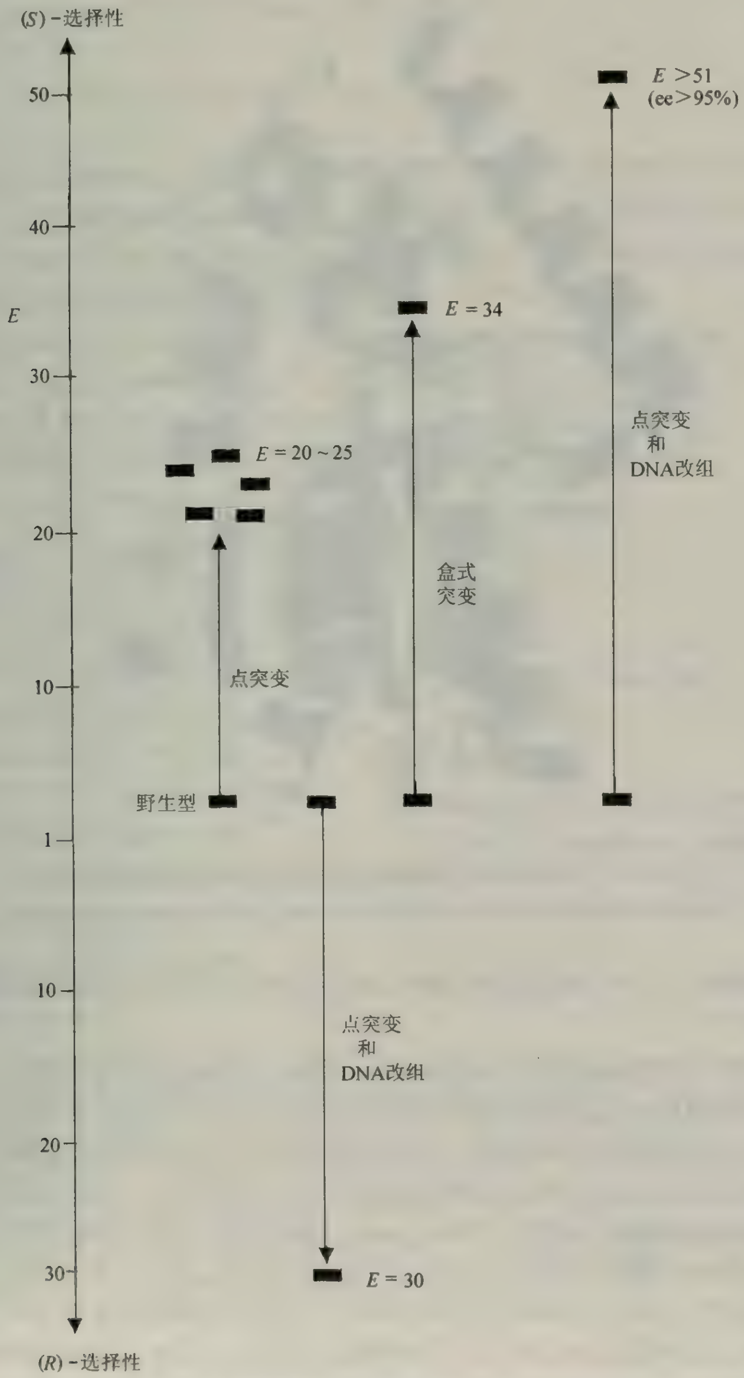


图 5.36　脂肪酶突变体催化酯 9(S)-或(R)-选择性

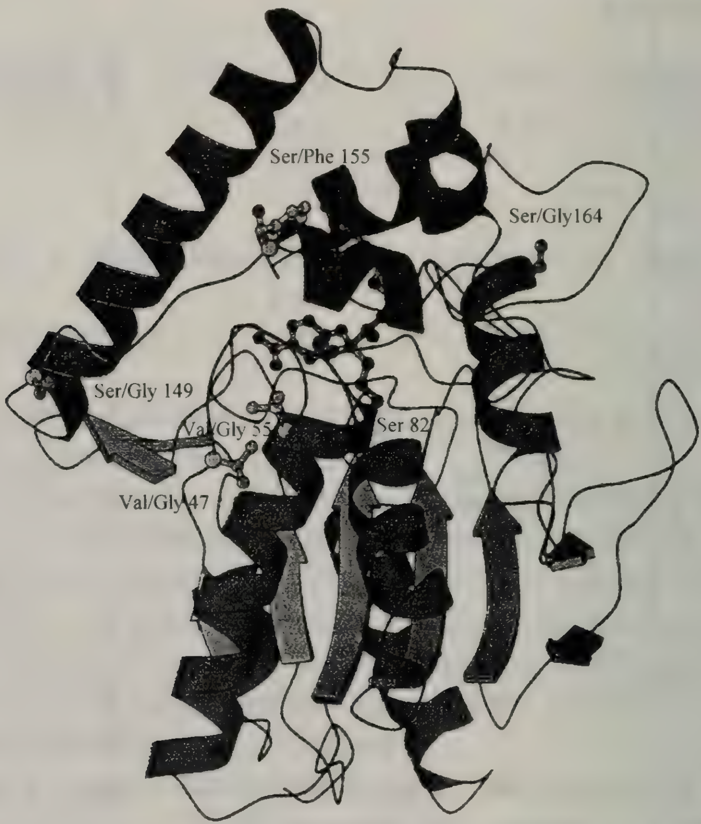


图 5.37　*Pseudomonas aeruginosa* 野生型脂肪酶 X 射线结晶结构

图中标出对映体选择突变体 94G12(V47G、V55G、S149G、S155F 和 S164G)及结合底物(*S*)-9

股 β4 和螺旋 α2 之间区域的柔性,同时,也可以影响在催化上主要的氧阴离子穴的稳定构象。

氨基酸 Ser149、Ser155 和 Ser164 定位在螺旋 α5 和 α7 之间的区域。它们的取代可以影响螺旋 α5 之后区域的次级结构,不仅仅因为增加柔性,而且也因为丧失 Ser149 和 Ser164 侧键之间的氢键。在突变酶 94G12 中,甚至可以形成螺旋 α5 对 α7 的各种相对定位,这个区域含有酶的底物结合部位各种残基。因此,这些螺旋的复位也影响酶的底物结合性质,特别是 Leu162 侧链的复位可能对酶的对映体选择性有很大的影响,可能直接同(*S*)-底物的甲基相互作用。此外,底物结合口袋残基 Leu50、Thr158、Leu62 同 Met16(涉及氧阴离子穴形成)的侧链行使范德华力相互作用。因此,这些残基位置的变化可以影响氧阴离子穴的结构,这样间接地影响酶的对映体选择性。取代的 S155F 明显的正效应是 Phe155 残基以范德华力距离带到

底物的酰基基团上,因此直接影响酶的对映体选择性。

5.3.8.3　脂肪酶最佳化的途径[35]

一些有效的脂肪酶(三酰基甘油水解酶;EC3.1.1.3)自 19 世纪 80 年代以来明显增加。这主要是由于微生物酶的基因克隆和表达取得了巨大的进步和日益增长的需要。在 20 世纪 80 年代中期以前,脂肪酶主要用于洗衣和甘油三酯的修饰。但广泛研究证明它们对光学纯化合物的合成也非常有效。它们有很好的稳定性,在广泛范围的有机溶剂中有活性,而且不需要辅因子。现在应用脂肪酶有效的拆分 1000 多种不同的物质。此外,以脂肪酶催化为基础的几种工业过程已经建立,根据产品的价值,BASF 每年生产 2000t 用于动力学拆分各种胺,这是工业化成功最突出的例子。然而,酶在工业过程中不总是有效的,因此应该根据需要进一步改进。表 5.1 指出了改进脂肪酶的策略,这些策略不仅适用于脂肪酶,而且也适用于其他酶。

表 5.1　改进脂肪酶的策略和方法

物理化学方法
固定化(改进稳定性)
增溶(溶于有机溶剂)
化学衍生(硝化作用)
反应工程方法
水解对酯化
水活度控制
酰基载体选择(乙烯基乙酸酯)
溶剂(离子液体)
分子生物学方法
合理蛋白质设计(改变对映体选择性)
定向进化(改变底物特异性,提高稳定性)

1. 物理化学方法

固定化是稳定生物催化剂并简化下游过程常用的物理化学方法。固定化脂肪酶和其他酶的一些载体和方法已发展,但是对于适宜材料和条件的选择仍没有普遍的准则,20 世纪 80 年代晚期和 90 年代早期已发展了几种非常稳定的固定化脂肪酶制剂,例如 Novozyme(Bagsvaerd, Denmark),Altus(Cambridge, MA, USA)等。在有机溶剂反应中,对脂肪酶的增溶作用进行了广泛的研究,但对不同溶剂的增溶没有普遍的方法,目前尚未见该研究领域重要成就的报道。

2. 反应工程方法

反应工程的应用可以在某种程度上改进某种脂肪酶的应用特性。在脂肪酶催

化反式酯化反应中应用不同的酰基供体,经常采用烯醇酯(例如乙烯基乙酸酯),因为它们不可逆并反应迅速。(疏水常数数据库)概念经常用于选择适当溶剂,但溶剂影响活性、稳定性和对映体选择性仍然是无法预测的。

最近的成就是发现酶催化反应可以在离子液体(Ils)中进行。Ils 最普通是 1-丁基-3-甲基咪唑鎓四氟硼酸盐([BMIM]BF_4 和六氟磷酸盐([BM1B]PF_6),而且,Ils 溶剂化性质可应用不同阳离子或阴离子调节,结果形成水-可溶混合水-不溶混系统。成功应用这些溶剂的第一个报道是发现南极洲假丝酵母(*Candida antarctica*)脂肪酶 B(CAL-B)的特性与在醇解、氨解和全水解中常规的有机溶剂相似。在 IL 存在下几个仲醇拆分时,CAL-B 和洋葱假单胞菌(*Pseudomonas cepacia*)脂肪酶(PCL)的对映体选择性可以保持或甚至提高。Kragl 等[36]详细讨论了离子液体中酶催化(表 5.2)。报道葡萄糖在 Ils 中乙酰化有更多地选择。然而,必须强调这些观察不是生物催化剂被 Ils 修饰的结果,而是溶剂系统极性改变的结果,大概是离子效应。

表 5.2 应用离子液体生物催化反应实例[35]

生物催化剂	IL	反应系统
E. coli 碱性磷酸酶	[H_3NEt][NO_3]	通过对硝基苯酚磷酸检验酶活性和稳定性
鸡卵清溶菌酶	[H_3NEt_3][NO_3]	蛋白质复性
红球菌	[BMIM][PF_6]/缓冲液(双相)	1,3-氰基苯生物转化
R312 细胞	[BMIM][PF_6]	红霉素提取
面包酵母细胞	[BMIM][PF_6]/缓冲液(双相)	酮还原
酵母细胞	[BMIM][PF_6]/缓冲液(双相)	Z-天冬甜素的合成
嗜热菌蛋白酶	[BMIM][PF_6]	由发酵液回收正丁醇的合成
α-胰凝乳蛋白酶	[OMIM][PF_6] [BMIM][PF_6]	*N*-乙酰-L-苯丙氨酸乙酯同 1-丙醇转酯反应
α-胰凝乳蛋白酶	[EMIM][BF_4]	*N*-乙酰-L-酪氨酸乙酯同 1-丙醇转酯反应
	[EMIM][$(CF_3SO_2)_2N$]	
	[BMIM][BF_4]	
	[BMIM][PF_6]	
	[MTOA][$(CF_3SO_2)_2N$]	

续表

生物催化剂	IL	反应系统
β-半乳糖苷酶	[BMIM][BF_4]/缓冲液(单相)	水解活性
枯草杆菌蛋白酶	[MMIM][$MeSO_4$]/缓冲液(单相)	*N*-乙酰乳糖胺合成
甲酸脱氢酶	[MMIM][$MeSO_4$]	NADH 再生
	[4-MBPy][BF_4]	
α-胰凝乳蛋白酶	[BMIM][$(CF_3SO_2)_2N$] [EMIM][$(CF_3SO_2)_2N$]	*N*-乙酰-L-苯丙氨酸乙酯同 1-丁醇转酯反应
CALB	[BMIM][PF_6]	醇解,氨解,全水解
	[BMIM][BF_4]	
筛选的 8 个脂肪酶和两个酯酶	10 种不同的 IL	(*R*,*S*)-1-苯乙醇动力学拆分
CALB,PCL	[EMIM][PF_6]	仲醇动力学拆分
	[BMIM][PF_6]	
CALB	[EMIM][BF_4]	通过转酯合成丁酰基丁酸
	[BMIM][$(CF_3SO_2)_2N$]	
	[BMIM][PF_6]	
	[EMIM][$(CF_3SO_2)_2N$]	
CALB,PCL,CRL,	[BMIM][PF_6]	烯丙基乙醇动力学拆分
猪肝脂肪酶	[BMIM][CF_3SO_3]	
	[BMIM][BF_4]	
	[BMIM][$(CF_3SO_2)_2N$]	
	[BMIM][SbF_6]	
CALB,PCL	几种 IL	(*R*,*S*)-1-苯乙醇动力学拆分;
	用碳酸钠水溶液洗	β-葡萄糖酰化
3 个脂肪酶	[BMIM][PF_6]	单酯的合成
	[BMIM][BF_4]	
假单胞菌脂肪酶	[BMIM][$(CF_3SO_2)_2N$]	(*R*,*S*)-1-苯乙醇动力学拆分;水活度和温度的影响
脂肪酶	[BMIM][PF_6]	在减压下酯基转移或 5-苯基-1-亚乙基六胺-3 醇动力学拆分
脂肪酶	[BMIM][PF_6] [BMIM][BF_4]	2-羟基甲基 1,4-苯丙二噁烷酯基转移

续表

生物催化剂	IL	反应系统
CALB	[EMIM][$(CF_3SO_2)_2N$] [BMIM][$(CF_3SO_2)_2N$]	在超临界 CO_2 存在下苯乙醇动力学拆分
CALB	[BMIM][$(CF_3SO_2)_2N$]	在超临界 CO_2 存在下 1-辛醇的酯化
脂肪酯,PFL AK	[BMIM][PF_6] [BMIM][BF_4]	P-手性羟甲基膦酸动力学拆分

BMIM, 1-butyl-3-methylimidazolium; CALB, Candida Antarctica lipase B; $(CF_3SO_2)_2N$, bis (trifluoromethyl) sulfonyl amide; CRL, Candida rugosa lipase; EMIM, 1-ethyl-3-methylimidazolium; MBPy, 1-butyl-4-methylpyridinium; MMIM, 1-methyl-3-methylimidazolium; MTBE, methyl tert-butylether; MTOA, methyl trioctylammonium; OMIM, 1-octyl-3-methylimidazolium; PCL, *Pseudomonas cepacia* lipase; PFL, *Pseudomonas fluorescens* lipase

脂肪酶在有机介质中的催化反应的另一个经常研究的参数是水活度(a_w)。水活度不仅仅在给定的溶剂中保持酶的催化活性,而且强烈影响酰化迁移有。这样,只有在小心控制水活度情况下才能获得高的产物收率。例如,脂肪酶催化由脂肪酸在 Sn1 位置交换合成磷脂,实际上是通过小心控制水活度实现的。

3. 分子生物学方法

(1)合理蛋白质设计

结晶学工作者已经解析许多脂肪酶的结构,计算机模型已用于这些酶的底物特异性和对映体选择性合理设计。了解脂肪酶的对映体选择性的起源的各种途径,以及通过合理设计改进活性和选择性已有诸多报道。

根据 Pleiss 等[37]工作表明,各种脂肪酶活性部位有不同的型主(motif)。当催化时,它负责四面体中间体的稳定化,这种型主称为 GX 或 GGGX 型主(氨基酸以单字母表示,X 表示任一氨基酸),并且定位在氧阴离子结合口袋中。到目前,型主仅以现象学方式进行了描述。某些研究表明该型主也控制脂肪酶对叔醇的活性。已知含有 GGGX 型主的圆柱状假丝酵母(*C. rugosa*)脂肪酶(CRL)既作用于叔醇的乙酸酯,也明显作用于一些其他的 GGGX 型主水解酶[3 个乙酰胆碱酯酶、猪肝酯酶和枯草杆菌(*Bacillus subtilis*)对硝基苄基酯酶(BsubpNBE)]水解 4 个模型乙酸酯(中等对映体选择性,$E = 1.5 \sim 11$)。根据精确的计算机模型预测各种突变体对 BsubpNBE 水解,表明在(R,S)-2-苯基紫铆黄酮-2-乙酸酯拆分中对映体选择性至少改进 6 倍。GX 型主水解酶对该模型化合物完全失活。

(2)定向进化

定向进化的主要挑战是由 $10^3 \sim 12^8$ 突变体库鉴定出所要求的突变体,或者通

过选择(即通过与生长相关的检验)或用高通量技术筛选。应该强调的是一种方法很难产生最佳化过程,需要各种方法的组合才能达到最佳效果。

5.3.8.4　乙内酰脲对映体选择性转化

E. coli 产生乙内酰脲酶对映体选择性转化,该途径与(*S*)-选择脂肪酶和(*R*)-选择脂肪酶的进化相似。所有已知的乙内酰脲酶都能对 D-5-(2-甲基硫代乙基)乙内酰脲(D-18)[导致 *N*-氨甲酰基-D-甲硫氨酸(D-19)积累]进行选择,在这种情况下通过应用消旋酶或 pH > 8 转化是完全可能的(图 5.38)。

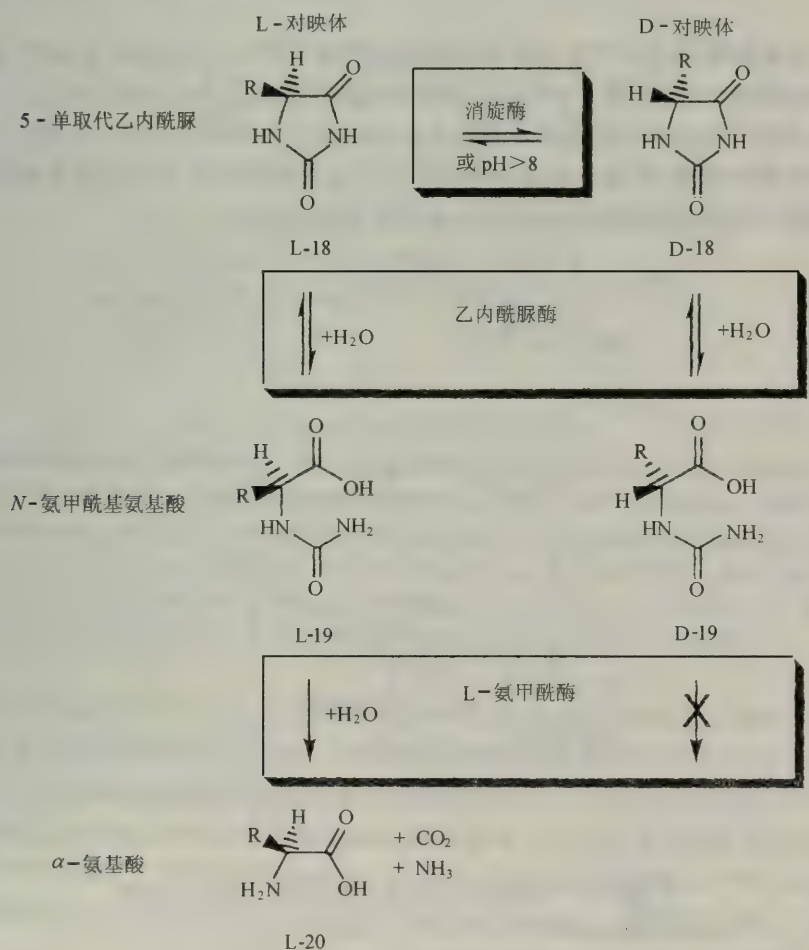


图 5.38　由消旋乙内酰脲生产 L-氨基酸所涉及的反应和酶

野生型乙内酰脲酶基因通过 epPCR 连续的几轮随机突变,继之饱和突变,所形成的库通过 pH 指示剂检验进行筛选[38]。携带 3 个氨基酸交换的突变体比野生型活力提高 1.5 倍,并且产生 *N*-氨甲酰基-L-甲硫氨酸(L-19),在 30%转化时,ee = 20%。当相应的基因同乙内酰脲消旋酶和 L-*N*-氨甲酰基酶在 *E. coli* 共表达时,导致前体化合物 D, L-5-(2-甲基硫代乙基)乙内酰脲转化 90%(2h 内与野生型 10h 反应结果相近)。虽然对映体选择性需要改进,这个结果明显证实通过单一酶基因的定向进化可以最佳化多酶途径,为改进细胞催化剂生产手性化合物提供了依据。

5.3.8.5 进化酶催化反应的立体化学

进化酶催化反应产生立体异构体有两种方式[39]:①进化酶,它接受非天然立体异构体作为底物[图 5.39(a)];②改变成键步骤的立体化学过程[图 5.39(b)]。这种方式可以由同一起始材料制备各种可能的立体异构体产物。目前已经用酶催化 P-O 键的形成,羰基还原烷烃的羟基化。定向进化改变 C-C 成键步骤的主体化学过程。该途径的成功对合成化学家具有重要的意义。

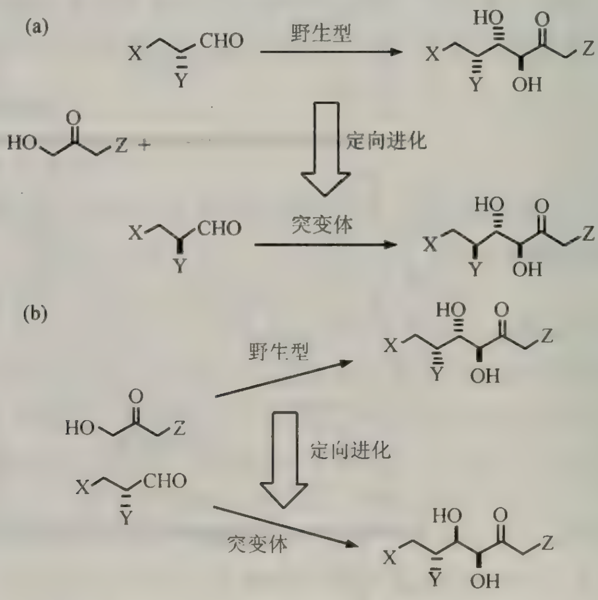


图 5.39 催化立体异构体产物形成的定向进化策略

(a)通过定向进化修饰底物的特异性;(b)通过定向进化修饰反应立体化学过程

醛缩酶(二磷酸果糖酶)催化 C-C 的合成,并用于复杂分子的合成。研究最好的是果糖 1,6-二磷酸(FBP)醛缩酶(EC4.1.2.13),这是普遍存在的糖解酶,其催化

二羟基丙酮磷酸(DHAP)和甘油醛 3-磷酸(G3P)缩合成果糖 1,6-二磷酸。醛缩酶可以根据它们的结构和反应机制分类:类 I 醛缩酶通常是$(\beta/\alpha)_8$桶亚基的多聚体,当催化时以 Schiff 碱形式利用活性部位的赖氨酸残基,而类 II 醛缩酶绝对需要Zn^{2+}。大肠杆菌的类 II FBP 醛缩酶也属于$(\beta/\alpha)_8$桶族。尽管对类 II FBP 醛缩酶的催化机制有了一定的了解,但关于立体化学控制的详节仍然不清楚。通过 Glu182 由 DHAP 提取 1-ProS 质子之后,形成的中间体进攻 FBP 醛缩酶中 G3P 的 Si-面产生 FBP[图 5.40(b)]。同样的中间体进攻 G3P 的 Re-面可能产生非对映异构体,塔格糖 1,6-二磷酸(TBP)[图 5.40(a)]。TBP 和 FBP 都是由 DHAP 和 G3P 合成的,它们仅仅在 C4 立体化学不同。因此,所合成产物的构型直接在 C-C 成键步骤产生。

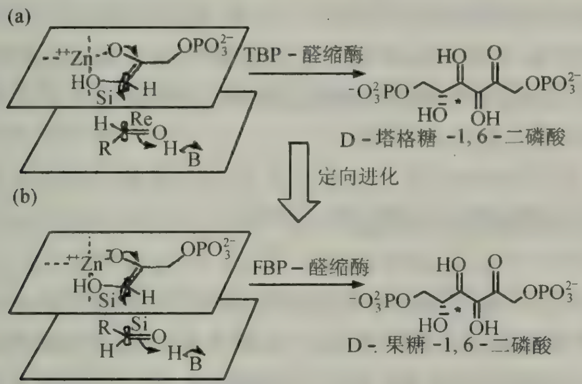


图 5.40　醛缩酶催化反应的立体化学

(a)TBP 醛缩酶的机制。DHAPene-diolate 是在由 DHAP 提取 1-ProS 质子之后形成,并通过催化锌阳离子极化。由它的 Si 面的活化的 DHAPC1 进攻 G3PC1 Re 面,产生 3S,4S 产物塔格糖 1,6-二磷酸,并且质子由 H-B(Asp-82)提供,将 C4 羰基转为羟基,竞争 TBP 合成。(b)FBP 醛缩酶的机制。中间体是在通过 Glu-182 由 DHAP 提取 1-ProS 质子之后形成的。由它的 Si 面活化的 DHAPC1 进攻 G3PC1 的 Si 面,并且质子由 H-B(Asp-109)提供,将 C4 羰基转为羟基,竞争 3S,4R 产物果糖 1,6-二磷酸的合成。FBP 和 TBP 在 C4 上是差向立体异构体(用 * 表示),$R = CH(OH)CH_2OP_3^{2-}$

反应的立体化学过程仅仅是受酶结合 FBP 和 TBP 过渡态相对自由能控制的。*E. coli* 类 II TBP 醛缩酶在分裂方向对 TBP 是高度特异的,高于它的非对映体 FBP 约 300 倍。类 II FBP 醛缩酶的顺序匹配和突变研究指出九个残基对底物的识别有潜在的影响,这些残基突变为 TBP 醛缩酶的相应残基。令人惊奇的是尽管 FBP 活性丧失,突变体没有产生 TBP 醛缩酶活性。这表明更复杂的取代组合产生 C-C 键形成的醛缩酶,把活性中间体加到底物醛的非对映异构面上,而且是可逆的。这就是说,定向进化可以用于改变塔格糖 1,6-二磷酸醛缩酶催化反应的立体化学过程。在三轮 DNA 改组和筛选之后,进化的醛缩酶对非天然底物果糖 1,6-二磷酸的 K_{cat}/K_m 提高 80 倍,立体选择性提高 100 倍,^{32}P NMR 光谱分析表明在合成方向

上,进化的醛缩酶催化具有非天然非对映异构体选择性的 C-C 键形成,使塔格糖 1,6-二磷酸转换为非对映异构体果糖 1,6-二磷酸。这证实定向进化对复杂的立体异构体产物合成的有效性。

定向进化已用于产生高活性,高稳定性,高底物选择性和高适应性的进化酶。通过定向进化也可以获得高对映体选性的酶,但需要克服两个主要的障碍:①检验对映体选择性高通量筛选系统;②探索对映体选择性的蛋白质顺序空间的各种策略的发展。虽然该领域进展迅速,但尚需更多的努力,包括根据选择(而不是筛选)ee 检验的方法,例如应用噬菌体展示或体外选择,探索蛋白质顺序空间方法、利用新底物把一种酶转为另一种新酶,例如把脂肪酶转为 Diels-Alderase 等。

遗传算法或神经网络的应用将有助于拓展蛋白质顺序空间,有助于对映体选择的酶的产生和发现,对映体选择酶的定向进化增加了酶功能开发的空间。这是由于对映体选择性是酶非常敏感的参数。目前,定向进化已成功地用于产生高对映体选择的酶和各种生物技术用酶。

5.3.9 $(\beta/\alpha)_8$-桶酶的功能进化[40]

$(\beta/\alpha)_8$ 桶是大约 10%结构表征的蛋白质共同的折叠型主(motif)。$(\beta/\alpha)_8$ 桶闭合的,平行的 β 折叠结构是由 8 个平行(β/α)单位通过氢键形成圆桶状核心。由于其八重伪对称,包装在桶内是 4 个$(\beta/\alpha)_2$ 亚结构域,其中各种疏水核心定位在 β 折叠和两侧 α 螺旋之间。活性部位定位在 β 股的 C 端。这样,围绕活性部位的功能团在结构上是独立的:“旧”和“新”酶在某些 β 股末端保留功能团,而其他可以改变以产生新的活性[36]。用此蓝图,$(\beta/\alpha)_8$ 桶是新功能进化的最佳候选者。随着顺序和结构数量的迅速增加,激励人们了解自然进化中如何利用$(\beta/\alpha)_8$ 桶骨架体外设计新反应的催化剂。

最近的研究指出,$(\beta/\alpha)_8$ 桶蛋白被分为 25 个超族,所有的$(\beta/\alpha)_8$ 桶的拓扑结构是相似的(闭合圆桶状结构),底物特异性的保持在进化过程中不是重要的。

$(\beta/\alpha)_8$ 亚结构域结构指出$(\beta/\alpha)_8$ 桶是由$(\beta/\alpha)_{2N}$前体组装的。分别有结合或催化作用的混合前体可以产生超族的原始酶。在这个过程后,功能的进化可以通过基因复制继之突变产生。功能进化的合理途径涉及编码原始酶基因的复制,这样当复制经历顺序和功能趋异进化时,可以保持原始功能。

原始酶和“新”酶所催化的反应之间的关系可以设想有以下 3 种策略。

1) 化学机制是优势。原始的机制可以用于催化“新”反应。最简单的例子,原始和新酶催化同样的化学反应,但底物特异性不同。更复杂的例子,各种反应共同使用结构策略,以稳定共同的中间体或过渡态。“烯醇酶 C 端类-结构域和依赖金属的水解酶”超族提供了这种策略的例子。

2) 配基特异性是优势。代谢途径可以“线性”进化(从起始到完成或反之亦然),这样原始酶和“新”酶具有共同底物/产物。His 和 Trp 生物合成中的各种酶是这种策略的例子。

3) 活性部位结构是优势。自然选择具有功能团的原始酶可以用于催化机制不同的“新”反应。这种策略的第一个例子是乳清酸核苷 5′-单磷酸(OMP)脱羧酶和 3-酮基-L-古洛糖酸 6-磷酸脱羧酶(KGPDC),机制不同的脱羧酶有同源顺序和结构。

具有$(\beta/\alpha)_8$ 桶折叠的烯醇酶超族、酰胺水解酶超族、乳清酸核苷单磷酸脱羧酶超族和醛缩酶的定向进化已有报道,主要集中在改变原始酶的底物特异性方面,少数是改变催化的化学反应。

(1) His 和 Trp 生物合成各种酶的进化

$(\beta/\alpha)_8$ 桶骨架定向进化最有价值的报道是 Fersht 等工作(见第 2 章第 4 节)。

(2) 烯醇酶超族的功能进化

烯醇酶超族中$(\beta/\alpha)_8$ 桶进化潜力也已探讨。*Pseudomonas* sp. P51 黏康酸内酯化酶(muconate lactonizing enzyme, MLEII), *E. coli* 邻-琥珀酰苯甲酸合酶(O-succinylbenzoate synthase, OSBS)和 *E. coli* L-Ala-D/L-Glu 差向异构酶在第二和第六 β 股末端含 Lys。在 MLE 中第二 β 股末端 Lys 启动质子传递到烯醇酸中间体上;在 OSBS 中,第二股末端 Lys 启动质子提取反应,并使留下的氢氧化物质子化;在差向异构酶中,Lys1-质子转移的双一碱机制,既不是 MLE 也不是差向异构酶催化 OSBS 反应。然而,由于这些酶是同源的,所有 3 个酶都是来自共同的原始酶,并且有限的取代就可以进行活性互换。差向异构酶和 MLE 的组合进化,两者产生催化 OSBS 反应的蛋白质,并且与它们的原始酶不同。在设计途径中,其中 *E. coli* 差向异构酶和 OSBS 的结构比较表明,差向异构酶的第八 β 股末端的同源天冬氨酸用甘氨取代,这样 OSBS 的更大的底物可以结合在活性部位。在 MLE 的组合途径中,OSBS 的活性来自第八 β 股末端的谷氨酸用甘氨酸同源取代。两个进化的蛋白质有明显的 OSBS 活性,并降低了原始酶的活性。在不同的原始酶中两种同源取代产生新活性,这表明$(\beta/\alpha)_8$ 桶折叠的功能可塑性。

(3) 酰胺水解酶超族功能进化

酰胺水解酶超族的特殊成员采取合理和组合突变,提高了底物特异性。例如,进化的磷酸三酯酶(PTE)对手性底物的固有立体选择性提高、松动和转化。PTE 的 DNA 改组改进了甲基对硫磷催化活性,比野生型酶增加 25 倍。

(4) 醛缩酶的功能进化

尽管在顺序和结构库中有大量的醛缩酶,目前它们功能的进化还仅仅局限于类 I(形成 Schiff 碱)、2-酮基-3-脱氧-6-磷酸葡糖酸(KDPG)醛缩酶。它是葡萄糖分解代谢 Enter-Doudoroff 途径中一种酶,产生丙酮酸和 D-甘油醛-3-磷酸。KDPG 醛

缩酶具有广谱底物特异性。

E. coli 醛缩酶已成功地进化，经易错 PCR(epPCR)和改组，活性是野生型 600 倍。改变活性中心 Lys 的突变体可以利用苯甲醛和电子等排吡啶羧基醛作为共底物(cosubstrate)，$(\beta/\alpha)_8$ 桶第六 β 股末端 Lys133 同底物形成 Schiff 碱。

$(\beta/\alpha)_8$ 桶酶是定向进化的亮点，从$(\beta/\alpha)_8$ 桶出发，催化新反应酶的设计和发现将是可行的，不久的将来或许是常规。

5.4　应用进化策略研究酶的结构和功能[1]

5.4.1　应用进化策略研究分支酸变位酶的结构和功能

利用定向进化策略研究酶的结构和功能，最典型的例子是分支酸变位酶(chorismate mutase)，它催化分支酸转化为预苯酸(图 5.41)。不同来源的变位酶虽然顺序相似性小，却表现出相似的动力学性质，但它们的三级和四级结构是不同

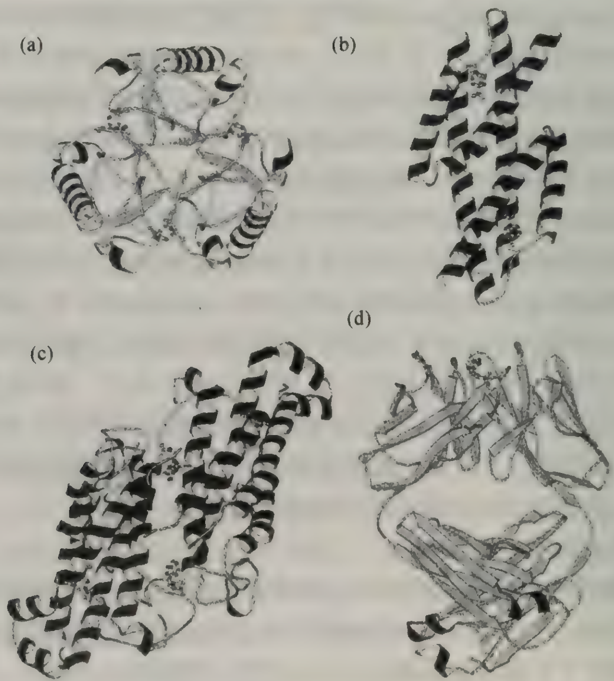


图 5.41　3 个天然分支酸变位酶和抗体酶的带状图解

图中黑色表示螺旋，灰色表示 β 折叠，在每个活性部位上的过渡态类似物抑制剂以球棒表示。(a) *B. subtilis* 的单功能分支酸变位酶(BcCM)；(b) *E. coli* 双功能分支酸变位酶-预苯酸脱水酶(EcCM)；(c)酵母 *S. cerevisiae* 别构调节分支酸变位酶(ScCM)。色氨酸(其能活化酶)结合在二聚体界面别构位点上。(d)抗体酶 1F7

的。例如 *Bacillus subtilis* 分支酸变位酶(BsCM),*E. coli*(EcCM)和酵母 *Saccharomyces cerevesiae*(ScCM)(图 5.41)。BsCM(AroH 类分支酸变位酶的成员)是一对称的均一二聚体,包装成伪 β/α 桶。它有 3 个相同的活性部位在亚基界面上[图 5.41(a)],相反,EcCM[图 5.41(b)]和 ScCM[图 5.41(c)]是与分支酸变位酶 AroQ 类的成员无关。顺序库的同源搜索发现其他的蛋白分属这些结构类型。分支酸变位酶活性的抗体酶具有典型的免疫球蛋白折叠,见图 5.41(d)。

E. coli KA12 是典型的细菌选择系统(图 5.42),该工程菌缺失双功能分支酸变位酶(分支酸变位酶-预苯酸脱氢酶和分支酸变位酶-预苯酸脱水酶)。质粒 pKIMP-UAUC 是携带缺失分支酸变位酶活性的预苯酸脱水酶和预苯酸脱氢酶的质粒。KA12/pKIMP-UAUC 细菌是不能在缺乏外源酪氨酸和苯丙氨酸情况下生长。

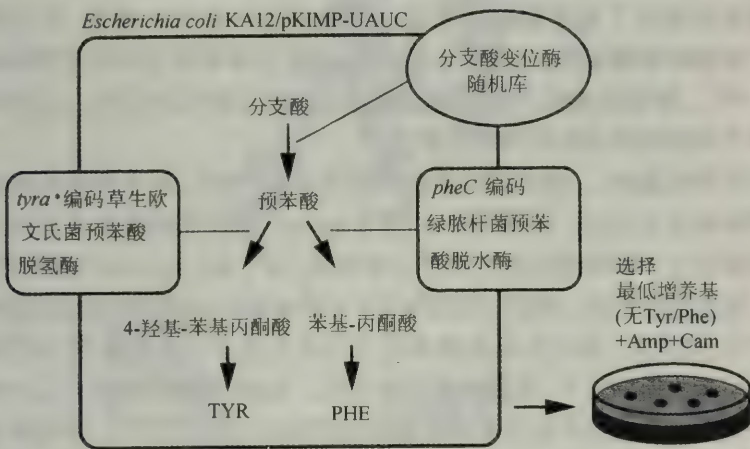


图 5.42　缺乏内源分支酸变位酶活性 *E. coli* 工程菌选择系统

编码双功能酶(分支酸变位酶-预苯酸脱氢酶和分支酸变位酶-预苯酸脱水酶)的基因缺失。脱氢酶和脱水酶的单功能突变体是由质粒 pKIMP-UAUC 提供。来自随机基因库的分支酸变位酶是根据它们能够补充遗传缺失进行估价,并允许细胞在缺乏 Tyr 和 Phe 的情况下生长。质粒库和 pKIMP-UAUC 分别携带 Amp^R 和 Cam^R 基因

BsCM 是由 127 个氨基酸残基组成的。前 115 个残基的结构已确定,末端 3_{10} 螺旋接近活性部位[图 5.41(a)]。一些实验表明柔性 C 端尾可能在酶功能上起作用。为了探讨这个问题,应用随机截短酶的 C 端,继之在 KA12/pKIMP-UAUC 系统选择功能克隆。用部分随机的寡核苷酸进行 PCR 操作,构建了两个基因库,其中相应于残基 116 ~ 117 或 111 ~ 127(包括 3_{10}螺旋)BsCM 密码子已随机化,产生随机截短的蛋白质。因为只有功能基因才能存活,编码蛋白质的必需部分可以迅速和直接的鉴定,这样可以发现决定结构或机制作用的残基。

在构建和转化到细菌选择菌株后,形成含有 10^4 ~ 10^5 成员的两个库。库中 C

端 12 残基突变体在选择介质中存活。这个结果证明 C 端残基不是酶活性所绝对需要的。甚至最后的 11 个残基可以去除而不明显影响体内酶活性。相反，当最后 17 个残基突变时，大约有 1/4 库能够在选择平板上生长。追加的 5 个随机氨基酸形成一个便于界定的 3_{10}螺旋，并且同蛋白质及其配基接触。没有发现短于 116 残基的活性酶。显然，3_{10}螺旋中的残基不能去除。几个选择的 BsCM 突变体的动力学特性指出，C 端尾的截短或突变很少影响酶的转换数（K_{cat}）。当分支酸被结合到突变体的活性部位时，其转化为预苯酸的效率与野生型 BsCM 接近。然而，K_{cat}/K_m 值实际降低是明显的。该发现指出 C 端，当不直接地涉及结合配体的化学转化时，通过底物和过渡态的均匀结合贡献给酶催化效率。

选择方法是检验影响蛋白质结构和稳定性的各种因素的理想方法。通过组合库和功能筛选探讨了决定蛋白质结构的螺旋间扭转作用。例如，突变细胞色素 b-562 3个残基螺旋间扭转，由随机库分离 31 个突变体取稳定的类天然结构。用突变 Rop（一种二聚结合 RNA 四螺旋束蛋白质）获得了相似的结果。这些结果指出对螺旋间片段的顺序或长度相对约束少。

催化活性的选择（一种结构完整性相当敏感的探针），较之简单的筛选方法可能有更微妙顺序优先性。应用 EcCM［图 5.41(b)］作为模板检验上述观点。两个相同的多肽由 3 个螺旋通过两个突环片段组成［图 5.43(a)］。L2 突环连接 H2 和 H3 螺旋，远离活性部位。最初，在扭转中 3 个暴露溶剂的残基（Ala65-His86-His67）［图 5.43(b)］随机化，通过体内选择鉴定催化活性突变体。实验指出扭转中的取代可能影响蛋白质的生产、稳定性或催化活性。Ala65-His86-His67 同 Lys64 随机组合，实验表明 64 位的随机化不影响残基在相邻扭转区的优先性。当埋藏的扭转残基 Leu68 同 Ala65-His86-His67 一起随机化，Leu68 的侧链埋藏在 L2 和 H1′（另一亚基的螺旋 H1）之间的界面，并且与溶剂分离（图 5.43）。顺序分析指出 68 位对疏水的脂肪族氨基酸有严格的要求。

现有蛋白质骨架的再工程代表从头设计功能酶的第一步。各种蛋白质寡聚态的改变，例如单体转化为多体，或多体转化为单体备受科学家关注。这样的系统可以弄清楚多体蛋白质的进化起源。

将寡体转化催化功能的单体是特别困难的，例如复杂的缠绕的均二聚体 EcCM。柔性的“铰链突环”插入到长 H1 螺旋（跨越 EcCM 二聚体）可以改变酶的四级结构。虽然 EcCM 活性部位是由两个多肽链所贡献的残基组成的，这样的片段可以允许 H1 N 端的一半自身回折，取代第二个肽，产生单体和催化活性的四螺旋束（图 5.44）。这取决于突环长度和组成。蛋白质的稳定性是寡聚蛋白质进化的一种重要驱动力。突环的长度似乎起更重要的结构作用。突环顺序可以改变四级结构而不影响活性部位结构是相当少的。

虽然 EcCM 均二聚体拓扑转化为单体保留酶活性部位，但产生不稳定的蛋白

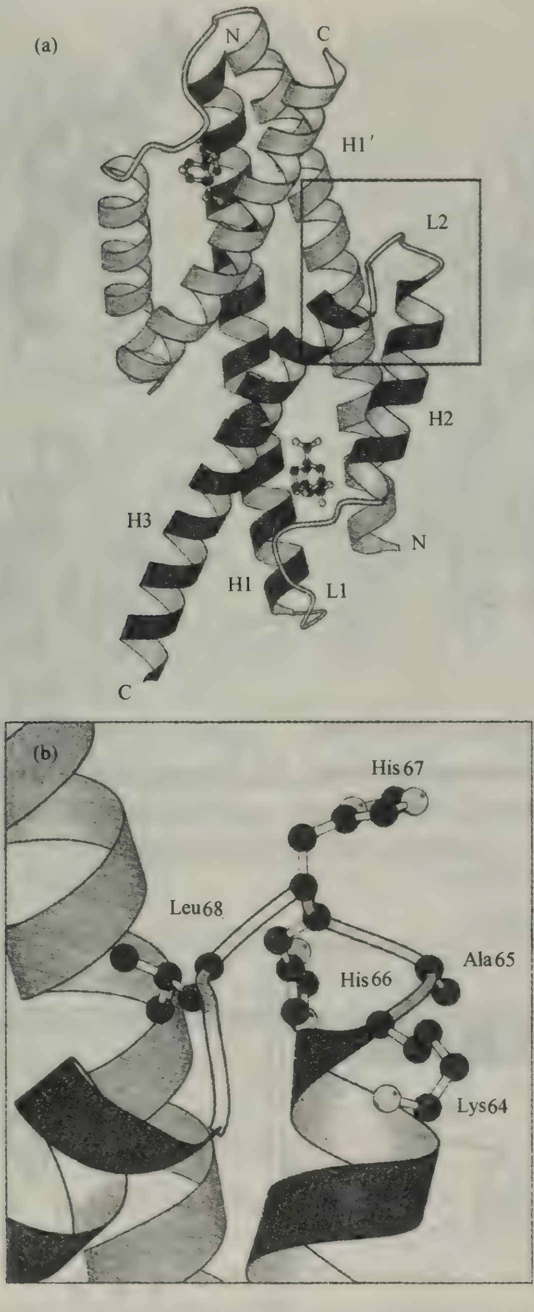


图 5.43　EcCM 的带状结构(a)和 L2 突环(b)

L2 的随机突变继之体内选择分支酸变位酶活性，结果表明对暴露溶剂的扭转残基顺序约束少，稍偏向疏水氨基酸。相反，常呈三级接触，68 位严格要求疏水的脂肪族氨基酸



图 5.44 二聚 EcCM 拓扑转化为单体

柔性突环(黑色)插入到跨越 H1 螺旋的二聚体中,允许螺旋的 N 端部分自身回折,取代其他多肽形成有功能的四螺旋束催化剂

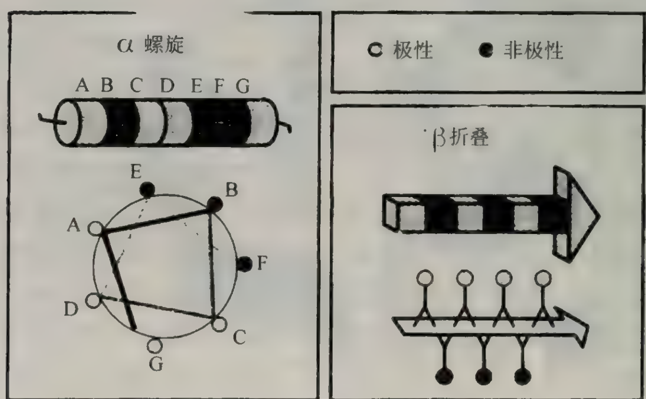


图 5.45 简单的双元模型

图中灰色表示极性氨基酸的位置;黑色表示非极性氨基酸的位置。据此模型可以设计两性螺旋或 β 折叠

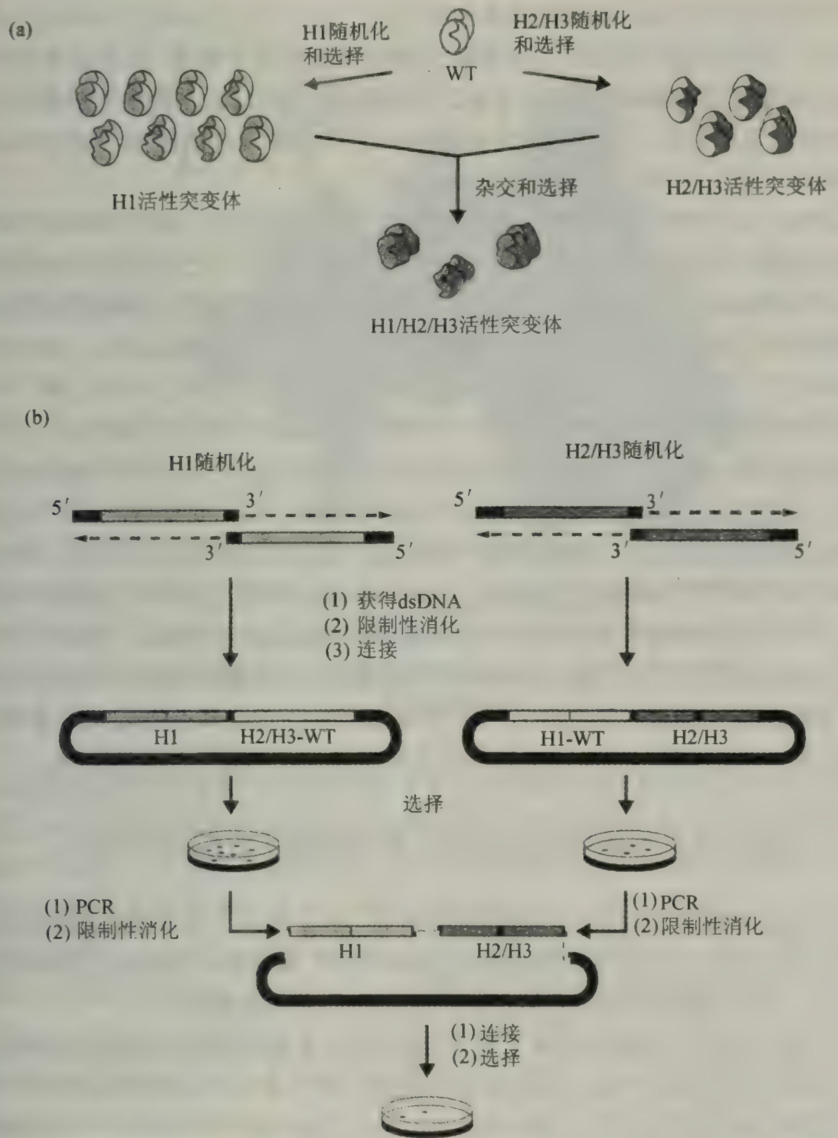


图 5.46　AroQ 分支酸变位酶基因双元模型库一般图解

(a) 双阶段策略。H1 螺旋（灰色）和 H2/H3 螺旋（黑色）分别随机化和选择功能突变体，继之初始库功能双元模型片段组合和再选择，最终构建物约 80%随机化。(b) 双元模型基因由编码 H1 或 H2/H3 螺旋片段的随机合成的寡核苷酸构成，并组装在载体片段上（黑色）。在初始步骤中，随机化片段同它们的野生型 H2/H3-WT 和 H1-WT（白色）组合。应用图 5.42 所示系统进行选择。WT-野生型 MjCM

质。两个活性部位之间的疏水残基簇似乎稳定二聚体,但这些相互作用在单体中可能丧失。这些埋藏的非极性基团暴露于溶剂大概有利于凝聚,因为高电荷结合口袋定位在螺旋束的内部,单体本身缺乏常规的疏水核心(这是蛋白质折叠的驱动力)。因此热稳定性的二聚体被认为是稳定单体变位酶设计的起点。通过消除失活或弱活性克隆已选择出稳定的单体变位酶。

从原理上讲,同样的进化途径可以成功用于蛋白质设计。极性和非极性氨基酸的双元模型可能是蛋白质设计的一般策略(图 5.45)。它已成功地产生许多天然蛋白质的螺旋束。例如,用随机顺序双元模型单元系统取代分支酸变位酶中的二级结构单元(图 5.46),然后用遗传选择估价库中蛋白质的催化活性。实验的第一步是相应于 H1 和 H2/H3 螺旋的模块个体同热稳定的 MjCM′(一种二聚产甲烷球菌分支酸变位酶的抗蛋白酶突变体)互补野生型螺旋片段组合(图 5.46)。在每个随机位置仅有 4 个极性氨基酸(Arg、Asp、Glu 和 Lys)或 4 个非极性(Ile、Leu、Met 和 Phe)。在初始设计时保持突环和保守的催化残基。双元模型模块库转化到分支酸变位酶缺陷菌株,并且能够使部分随机化的蛋白质折叠成有序结构。催化活性结构用分支酸变位酶缺陷互补估价,由 H1 和 H2/H3 库已分离出活性酶并已进行表征,所有的活性酶在 84 位和 85 位严格保守 Asn-Lys 二元组。预选的 H1 和 H2/H3 模块组合交叉,产生 H1/H2/H3 双元模型库,活性酶约占 1/10 000[图 5.46(a)]。选择的 H1/H2/H3 突变体之一的生物化学表征确定其是螺旋均二聚体,它催化分支酸转为预苯酸,这清楚地表明双元模型是催化剂设计的有效途径。

5.4.2 应用进化策略研究 DNA 聚合酶的结构和功能

在一定的条件下,每种蛋白质顺序将保持一个适应度。每种蛋白质的适应度在顺序空间中作图,三维景观形成适应度的峰和谷,最高峰代表最适合的蛋白质(图 5.47)。因此,酶的成功改进可以看作通过突变达到最高峰。

DNA 聚合酶定向进化的主要目标是提供 DNA 复制时高度忠实的结构单元。

聚合的忠实性取决于聚合酶活性部位能够分辨互补和非互补掺入的核苷三磷酸以及延伸匹配和非匹配的引物末端。Taq DNA 聚合酶不含有 3′-5′外切核酸酶或校正活性,这使其成为研究聚合酶活性部位内在忠实性的目标。在催化时,聚合酶至少经历 4 种明显的构象变化:①与 DNA 结合;②dNTP 结合;③化学催化,构象转换,形成磷酸二酯键并释放 PPi;④引物 3′羟基末端转移。理论研究指出多重构象改变贡献给 DNA 聚合酶催化的忠实性。DNA 聚合酶有 A、B 和 C 型主,前两者有 13 个氨基酸残基,C 型主有 11 个氨基酸残基。在型主 B 的 13 个氨基酸中,用随机突变产生 5×10^4 转化子库,由该库筛选出 67 个活性突变体,其中 13 个有延伸活性。两个突变体(A664R 和 A661E)已纯化,正向突变频率提高 7 ~ 25 倍。这两

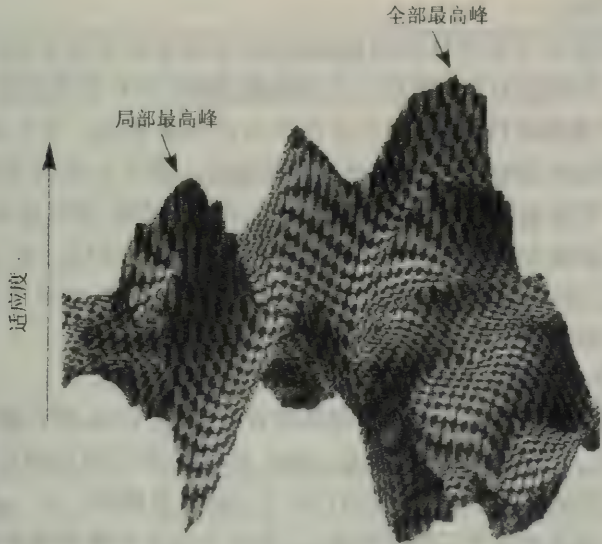


图 5.47　顺序空间的图解

这种景观表明一种酶可能的氨基酸顺序。峰高表示在某种条件下酶的适应度。单一氨基酸改变有局部最高峰;一组同时改变可以搜索到全部最高峰

个残基接近"指状"亚结构域的末端,并且可以稳定 dNTP 结合之后的闭合构象,这样使 DNA 合成的忠实性降低。已经筛选出高保真的突变体,三重突变体(A661E, I665T,F667L),其与野生型酶比较,突变频率降低 3 倍。该突变体提高了对 A:A 错配的分辨,并降低错配延伸。A661E 是强突变子,I665T 或 F667L 突变作为基因内校正基因。由型主 A 随机化获得了高和低忠实性的突变体,低忠实性突变体在 I164 位有亲水性取代。这个位置是能够耐受各种取代,但仍能保持野生型酶活性水平。当用于 PCR 反应时,在 I614 位含亲水残基的 Taq pol1 突变体与野生型酶比较,错误速率提高 20 倍,它有效的催化转换和颠换。除了提供低碱基配对忠实性外,I614 位的亲水取代也允许突变体 Taq pol1 避开阻断模板损伤。I164 位的亲水取代大概通过拓宽 dNTP 结合口袋以容纳异常碱基和核苷酸而降低忠实性。这样的突变体将促进 PCR 方法的应用,以产生各种各样的突变基因。

就 Taq pol1 型主 A 和 B 随机化库而言,各种取代谱是允许的,不丧失催化活性。型主 A,只有 13 个残基之一(Asp610)是绝对不可改变,而型主 B,13 个残基中有两个(Arg659 和 Lys563)是不可改变的。这些残基是活性必需的。总之,随机突变资料表明,型主 A 和 B 中活性部位 26 个残基中 3 个是体外功能绝对需要的。6 个氨基酸仅仅耐受保守取代,所有其他的残基是高度可变的。这表明 Taq 聚合酶

活性部位的易变性。

E. coli pol1 型主 A 突变体的随机库中，所有残基都是可变的，突变谱与 Taq pol1 明显相似。特别是 D705(催化上主要的氨基酸)没有出现在活性突变体中。E710 只能由天冬氨酸取代，这表明在该位负电荷的重要性。型主 A 的 N 端和 C 端的单一取代倾向对 DNA 聚合酶活性有相反效应。型主 A 的 N 端突变产生的酶活性等于或好于野生型酶。该区域形成反平行β折叠结构，容纳底物 dNTP 的三磷酸部位。C 端区的突变倾向给出比野生型酶活性低的突变体。在 *E. coli* pol1 和 Taq pol1 之间的主要差别是 Taq 中大量的正电荷取代，并在 Y611 位限于平面环状氨基酸。这些差别可能是由于不同折叠和包装约束所致。

关于 DNA 聚合酶突变为有效的 RNA 聚合酶见图 3.22。

核糖核苷酸在细胞内浓度远高于脱氧核苷酸，因此 DNA 聚合酶专一地分辨掺入的核糖核苷酸。根据 Taq pol1 结合 dNTP 的活性部位的晶体结构，分辨涉及 Glu615 的氨基酸侧链，其提供移入 rNTP 的 2′-OH 立体障碍。*E. coli* pol1 中丙氨酸定位突变结果提高单核苷三磷酸的掺入，但突变体没有聚合酶活性。大概由于中性残基的引入减少了总的酶活性。在型主 A 随机突变后选择活性酶，然后筛选能够掺入 rNTP 突变体，分离出能够有效合成长 RNA 的 Taq 聚合酶。这导致提出一种假说，I164 取代，特别是疏水取代，导致更开放的活性部位，更易容纳改变结构的核苷酸。Ala597 和 Glu615 是绝对保守的。

自然界是通过达尔文进化机制进行蛋白质的设计，细胞中每种蛋白质，从酶和受体到结构蛋白质都是通过这种机制产生的。本节以分支酸变位酶和 DNA 聚合酶为例，说明进化策略可以成功地研究现有蛋白质的结构和功能，并可以创造出新的蛋白质。表 5.3 列出进化酶实例。

表 5.3 进化酶实例[11]

靶 酶	进化的性质	变 化
酰胺酶(*B. stearothermophilus*)	提高在 *E. coli* 中表达	23 倍
氨酰基-tRNA 合成酶	提高互不相关的校正 tRNA 的氨酰化	55 倍
砷酸盐解毒途径	提高活性	40 倍
天冬氨酸氨基转移酶	提高对β分支氨基酸和 2-氧代酸的活性；	10^5 倍
	提高对 Val 的活性	2.1×10^6 倍
联苯二氧酶	提高多氯代联苯(PCBs)的活性	降解各种 PCBs、多氯代苯和甲苯
β-糖苷酶 A	65℃稳定	$t_{1/2}$ = 12min

续表

靶　酶	进化的性质	变　化
β-糖苷酶的 celB (pyrococcusfuriosus)	提高 20℃时活性	3 倍
β-glucoronidase	在戊二醛处理后仍保持功能	对戊二醛和甲醛的稳定性提高
β-内酰胺酶	提高对 cefotasime 活性	32 000 倍
$(\beta\alpha)_8$-酶 IGPs[1)]	转换为 PRAI[2)]活性	PRAI 活性高于野生型酶 6 倍
$(\beta\alpha)_8$-酶 ProFARI[1)]	有 PRAI 活性同时保留 ProFARI 活性	PRAI 活性低于野生型 $3 \sim 11 \times 10^4$ 倍；ProFARI 活性低于野生型 26 倍
羧甲基纤维素酶	提高活性	2.2～5 倍
过氧化氢酶 1	提高内在的过氧化物酶活性	由 2%提高到 58%
儿茶酚-2,3-双氧酶	提高热稳定性	在 50℃活性提高 13～26 倍
头孢菌素酶	提高对 moxalactam 活性	270～540 倍
分支酸变位酶(二聚体)	变为可溶性单聚体形式	单体和六聚体酶活性
细胞色素 P450BM-3	改变底物专一性	吲哚的羟基化
细胞色素 P450 单氧酶	在没有辅因子的情况下提高对萘的活性	5～20 倍
细胞色素 C 过氧化物酶	提高对愈创木酚的活性	300 倍
D-选择的乙内酰脲酶 (*Arthrobacter* sp.)	对映体选择性和提高活性	转为 L-乙内酰脲酶，活性提高 5 倍
DNA 聚合酶 β	增加自发突变频率	10～30 倍
脱氧核苷酸激酶 (*Drosophila melamogaster*)	对核苷类似物敏感性	在 AZT[2)]存在下转化的 *E. coli* LD_{100}降低 316 倍；对 ddC[2)]降低 11 倍
酯酶(*Pseudomonas fluorescens*)	提高对映体选择性	2 倍
	空间障碍 3-羟基酯的水解	对映体活性提高(0%到 25%ee)
Flp 重组酶	提高热稳定性	改进重组效率
真菌过氧化物酶	提高热和氧化稳定性	热稳定提高 174 倍，氧化稳定性提高 100 倍
半乳糖苷酶	转为岩藻糖苷酶	对对硝基苯呋喃糖苷的专一性提高 1000 倍，比活性提高 66 倍
谷胱甘肽转移酶	底物专一性	对新的各种底物的活性

续表

靶 酶	进化的性质	变 化
HIV 逆转录酶	对核苷类似物的抗性	依赖 DNA 和 DNA 聚合酶的活性和对 AZT 的抗性
马心脏肌红蛋白	提高内在的过氧化物酶活性	25 倍
辣根过氧化物酶	提高对 ABTS[2] 和愈创木酚的活性	总活性提高 40 倍；对 ABTS 活性提高 5.4 倍；对愈创木酶提高 2.3 倍
3-异丙基苹果酸脱氢酶（*B. subtiis*）	提高热稳定性	在 70℃活性提高 3.4 倍
卡那霉素核苷酰转移酶（*B. stearothermophilus*）	提高热稳定性	在 60～65℃的 $t_{1/2}$ 增加 200 倍； 提高 20℃
乳酸脱氢酶（*B. strarothermophilus*）	在没有辅因子果糖 1,6-二磷酸的情况下提高活性	70 倍
脂肪酶（*Pseudomonasaeruginosa*）	提高对映体选择性	在对映体选择性提高（2% 到 81%，ee）
脂肪酶（*Staphylococcushyicus*，*s. aureus*）	底物专一性（磷脂对短链脂肪酸酯）	对长链 pNB[2] 酯活性提高 3 倍
脂肪酶（*S. aureus*）	底物专一性（对磷脂的活性）	磷脂肪酶绝对活性提高 11.6 倍，磷脂肪酶/脂肪酶比例增加 11.5 倍
N-氨基甲酰酶-D-乙内酰脲酶融合蛋白	融合蛋白的稳定性	D-氨基酸的产量增加 6 倍
O^6-烷基鸟嘌呤烷基转移酶	对抑制剂 BG[2] 和 *N*-甲基-*N'*-硝基-*N*-亚硝基胍的抗性	BG 的抑制浓度减少到 50%
O^6-甲基鸟嘌呤甲基转移酶	对 *N*-甲基-*N'*-硝基-*N*-亚硝基胍体内抗性	突变频率降低 2.7～5.5 倍
八氢番茄红素脱氢酶，番茄红素环化酶	新的类胡萝卜素途径	合成 3,4,3′,4′-四氢类胡萝卜素和 torulene
pNB 酯酶	提高热稳定性	在所有温度下活性没有任何降低，T_m 增加 14℃

续表

靶　酶	进化的性质	变　化
pNB 酯酶	提高在水/有机溶剂中的活性	在25%～30% DMF 中对各种 pNB 酯的活性提高50～150倍
蛋白质水解酶 SS Ⅱ (*B. sphaericus*)	提高10℃时的活性	活性提高6倍,在70℃活性降低3.3倍
限制性内切核酸酶 *Eco*RV	延伸识别位点	识别8bp 和10bp 位点
丝氨酸蛋白质水解酶(*Lysobacter enzymogenes*)	改变专一性和提高活性	速率提高20～45倍,选择性提高
枯草杆菌蛋白酶(*B. lentus*)	分泌酶的表达水平	增加50%
各种枯草菌蛋白酶	各种性质	活性、稳定性提高
枯草杆菌蛋白酶 BPN′	在降低温度时活性提高	在10℃时活性提高2倍
枯草杆菌蛋白酶 E	水/有机溶剂中的活性	在60% DMF 中活性提高170倍
	提高热稳定性	最适温度提高17℃,在所有温度下活性均提高
	提高热稳定性	在65℃,$t_{1/2}$增加50倍
枯草杆菌蛋白酶 S41 (*Psychrophilic*)	提高热稳定性	$t_{1/2}$增加100倍
Taq DNA 聚合酶 I	在体内 *E. coli* DNA 聚合酶 I 的功能互补	活性部位突变体的活性保留
	提高突变频率	7～25倍
T7 RNA 聚合酶	增加突变频率	≥20倍
胸腺嘧啶脱氧核苷激酶 (HSV)[1]	提高对 ganciclovir 或 aciclovir 敏感性	对 ganciclovir 敏感性提高43倍,对 aciclovir 敏感性提高20倍
胸腺嘧啶脱氧核苷激酶-1或-2 (HSV)	降低对 AZT 的敏感性	TK-1,32倍;TK-2,16 000倍

1）见 http://www.che.caltech.edu:80/groups/fha/Enzyme/directed.html.

2）缩写:

ABTS = 2,2′-azinobis-(3-ethylbenzothiazoline-6-sulfonic) acid;

AZT = 3′-azido-3′deoxythymidine; PRAI = phosphoribosyl anthranilate isomerase;

IGPS = indoleglycerol phosphate synthase; BG = O^6-benzylguanine; ddC = dideoxycytidine; pNB = *para*-nitrobenzoate;

ProFARI = *N*′-[(5′-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide isomerase.

参 考 文 献

1 Brakmann S, Johnsson K. Directed molecular evolution of proteins. WILEY-VCH, 2002

2 Lynch M. Gene duplication and Evolution. Science, 2002, 297: 945 ~ 947

3 Schlunegger M P, Bennett M J, Eisenberg D. Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. Adv Protein Chem, 1997, 50: 61 ~ 122

4 Ostermeier M, Benkovic S J. Evolution of protein function by domain swapping. Adv Protein Chem, 2000, 55: 29 ~ 77

5 Ganen B. In comprehensive natural products chemistry, vo15: enzymes, enzyme mechanisms, proteins and aspects of NO chemistry (Ed.: Pouler C D). UK: Elsevier, Oxford, 1999, 343

6 Patthy L. Protein evolution. MA: Blackwell Science, Malden, 1999

7 Zhang H Y, Li Z Q, Zhang J. Enzymatic generation of libraries *in vitro* for random mutagenesis of the aspartase gene. Chinese Science Bulletin, 1992, 37: 598 ~ 601

8 Zhang H Y, Zhang J, Lin L, *et al*. Enhancement of stability and activity of aspartase by random and site-directed mutagenesis. Biochem Biophys Res Commun, 1993, 192: 15 ~ 21

9 Håkansson K, Wang, AH-J Miller C G. The structure of aspartyl dipeptidase reveals a unique fold with a Ser-His-Glu catalytic triad. Proc Natl Acad Sci USA, 2000, 97: 14097 ~ 14102

10 Kong X D, Liu Y M, Zhang J, *et al*. Directed evolution of α-aspartyl dipeptidase from Salmonella typhimurium. Biochem Biophys Res Commun, 2001, 280: 396 ~ 400

11 张今等. 分子酶学工程导论. 北京:科学出版社,2003

12 Kong X D, Li Z Q, Zhang J, *et al*. A monomeric L-aspartase obtained by *in vitro* selection. J Biol Chem, 2002, 277: 24289 ~ 24293

13 明镇寰,潘建伟,朱睦元. 非核糖体多肽合成酶研究进展. 生物化学与生物物理进展,2002,29:667 ~ 669

14 Kohli R M, Takagi J, Walsh C T. The thioesterase domain from a nonribosomal peptide Synthetase as a cyclization catalyst for integrin binding peptides. Proc Natl Acad Sci USA, 2002, 99: 1247 ~ 1252

15 Tsuji S Y, Wu N, Khosla C. Intermodular communication in polyketide synthases: comparing the role of protein-protein interactions to those in other multidomain proteins. Biochemistry, 2001, 40: 2317 ~ 2325

16 Nixon A E, Benkovic S J. Improvement in the efficiency of formyl transfer of a GAR transformylase hybrid enzyme. Protein Eng, 2003, 13: 323 ~ 327

17 Kong X D, Liu Y, Zhang J, *et al*. Directed Evolution of Operon of Trehalose-6-phosphate Synthase/Phosphatase from *Eschericia coli*. Biochem Biophys Res Commun, 2001, 280: 396 ~ 400

18 Schiraldi C, Lernia I D, Rosa M D. Trehalose production: exploiting novel approaches. Trends in Biotechnolgy, 2002, 20: 420 ~ 425

19 Mivazaki K, Wintrode P L, Gravling R A, *et al*. Directed evolution study of temperature adaptation in a psychrophilic enzyme. J Mol Biol, 2000, 297: 1015 ~ 1026

20 Fontana A *et al*. Rigidity of thermophilic enzymes. In: Ballesteros A, *et al*. Stability and Stabilization of Biocatalysts. Elsevier, Science, 1998. 277 ~ 294

21 Tamakoshi M, Yamagishi A, Oshima T. Screening of stable proteins in an extreme thermophile, Thermus thermophilus. Mol Microbiol, 1995, 16: 1031 ~ 1036

22 Meijer P J, Holmberg N, Grundstrom G, *et al*. Direct evolution of a type I antifreeze protein expressed in *Escherichia coli* with sodium chloride as selective pressure and its effect on antifreeze tolerance. Protein Eng, 1996,

9: 1051 ~ 1054

23 Kano H, Taguchi S, Momose H. Cold adaptation of a mesophilic serine protease, subtilisin, by *in vitro* random mutagenesis. Appl Microbiol Biotechnol, 1997, 47: 46 ~ 51

24 Strausberg S L, Alexander P A, Gallagher D T. Directed evolution of a subtilisin with calcium-independent stability. Biotechnology, 1995, 13: 669 ~ 675

25 Bessler C, Schmitt S, Maurer K H, *et al*. Directed evolution of a bacterial α-amylase: Toward enhanced pH-performance and higher specific activity. Protein Science, 2003, 12: 2141 ~ 2149

26 Schmidt-Dannert C. Directed Evolution of single proteins, metabolic pathways and viruses. Biochemistry, 2001, 40: 13125 ~ 13136

27 Gershenson A, Schauerte J A, Giver L, *et al*. Tryptophan phosphorescence study of enzyme flexibility and unfolding in laboratory-evolved thermostable esterases. Biochemistry, 2000, 39: 4658 ~ 4665

28 Altamirano M M, Blackburn J M, Aguayo C, *et al*. Directed evolution of new eatalytic activity using the α/β-barrel scaffold. Nature, 2000, 403: 617 ~ 622

29 Zhang J H, Dawes G, Stemmer W P. Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. Proc Natl Acad Sci USA, 1997, 94: 4504 ~ 4509

30 Leemhuis H, Rozeboom H J, Wilbrink M, *et al*. Conversion of Cyclodextrin glycosyltransferase into a starch hydrolase by directed evolution: The role of alanine 230 in acceptor subsitetl. Biochemistry, 2003, 42: 7518 ~ 7526

31 Fong S, Machajewski T D, Mak C C, *et al*. Directed evolution of D-2-Keto-3- deoxy- 6-phosphogluconate aldolase to new variants for the efficient synthesis of D-and L-sugars. Chem Biol, 2000, 7: 873 ~ 883

32 Jaeger K E, Eggert T, Eipper A, *et al*. Directed evolution and the creation of enantioselective biocatalysts. Appl Microbiol Biotechnol, 2001, 55: 519 ~ 530

33 Arnold F H. Combinatorial and computational challenges for biocatalyst design. Nature, 2001, 409: 253 ~ 257

34 Nardini M, Lang D A, Liebeton K, *et al*. Crystal structure of pseudomonas aeruginosa lipase in the open conformation. The prototype for family I.1 of bacterial lipases. J Biol Chem, 2000, 275: 31219 ~ 31225

35 May O, Nguyen P T, Arnold F H. Inverting enantioselectivity by directed evolution of hydantoinase for improved production of L-methionine. Nat Biotechnol, 2000, 18: 317 ~ 320

36 Bornscheuer U T, Bessler C, Srinivas R, *et al*. Optimizing lipases and related enzymes for efficient application. Trends in Biotechnology, 2002, 20: 433 ~ 437

37 Krag U, Eckstein M, Kaftzik N. Enzyme catalysis in ionic liguids. Current opinion in Biotechnology, 2002, 13: 565 ~ 571

38 Pleiss J, Fischer M, Peiker M, *et al*. Lipase engineering database: Understanding and exploiting sequence-structure- function relationships. J Mol Catal B Enzym, 2000, 10: 491 ~ 508

39 Williams G J, Domann S, Nelson A, *et al*. Modifying the stereochemistry of an enzyme- catalyzed reaction by directed evolution. Proc Natl Acad Sci USA, 2003, 18: 3143 ~ 3148

40 Gerlt J A, Raushel F M. Evolution of function in $(\beta/\alpha)_8$-barrel enzymes. Current opinion in chemical biology, 2003, 7: 252 ~ 264

第 6 章　外显子改组定向进化酶

6.1 引　　言

基因进化的各种机制和过程，直至 20 世纪 70 年代后期和 80 年代初期还是不清楚的。20 世纪 90 年代初两个有利的条件促进了基因进化机制和过程的研究，一是 DNA 和蛋白质顺序和结构数据库的发展；二是新基因概念的引入。目前，数据库储存的顺序和结构信息已达天文数字，使人们能够比较各种基因结构。基因组顺序和新的分子技术改变了我们关于基因和基因组组成的观点。

在断裂基因（split gene）发现不久，人们就认识到内含子的存在可以深刻地影响蛋白质的进化。Gilbert[1]指出内含子内的重组可以把外显子分开，内含子的中间重复顺序可能是改组外显子顺序的热点。Blake[2]支持这样的进化机制，他提出外显子是编码蛋白质的功能单元。

大多数真核生物结构基因中存在内含子，除了发根农杆菌（*Agrobacterium rhizogenes*）质粒、*E. coli* T2、T4 和 T6 噬菌体及几种枯草芽孢杆菌（*Bacillus subtilus*）HMU 噬菌体之外，原核生物则缺乏内含子。对此解释有两种假说：一种是后生内含子（intron late）假说。它指出原核生物基因类似祖先基因，而内含子来自易位单元，后来插入真核生物基因中。事实上，真核生物基因的外显子-内含子结构不是静止的，是内含子不断地插入和消除的动态过程。与此观点相反，早生内含子（intron early）假说认为蛋白质的基因中存在内含子反映了这些基因是由片段组装的，真核生物基因组的“基因片段”组织是原始形式。根据上述假说，真核生物保留内含子，增加了遗传的可塑性；另一方面，细菌通过消除内含子而增加功效。

内含子起源是科学家们经常讨论的课题。内含子出现很早，并用于组装第一个基因；内含子出现较晚，并像插入单元一样后来加上的，中断以前连续的基因。Gilbert[3]分析了中心模块和古代内含子进而提出内含子起源的混合理论：某些内含子是很古老的，并且用于原始外显子改组，而许多内含子是后来丧失和加上的。

关于大多数内含子是插入到真核生物基因组还是内含子早已存在于原核生物和真核生物的祖先基因中，还是某些内含子是很古老的，而许多内含子是后来丧失和加上的，仍然是一个争论的问题。

内含子本身也在进化，外显子改组与内含子进化是平行的。适合外显子改组的内含子出现在进化的相对后期阶段，因此，外显子改组可能在古代蛋白质构建中不起重要作用。

必须强调,通过外显子改组所产生的大多数模块蛋白质对后生动物的多细胞性是绝对重要的。例如,细胞外间质成分的出现是与第一个多细胞动物的出现紧密相关的。细胞外间质成分,膜结合蛋白质调节细胞-细胞和细胞-间质相互作用,受体蛋白质调节细胞-细胞通讯(例如受体酪氨酸激酶、受体酪氨酸磷酸酶)允许细胞以完整的方式起作用。事实上,细胞外间质的主要成分细胞黏连蛋白、受体蛋白都是由模块构建的。

20 世纪 70 年代晚期证实外显子改组是一种产生新基因的有效过程。各种基因的外显子新的组合,可以由非同源重组产生并进化成具有新功能的基因,内含子将加速新组合和新蛋白质进化的速率[4]。蛋白质数据库研究结果表明外显子改组是新基因起源的普通机制。

真核生物基因的内含子-外显子(intron-exon)组织指出外显子新的重组可以由插入的内含子顺序内的重组产生,从而获得具有功能改变的重排基因,基因组进化的这种模型称为外显子改组(exon-shuffling)[2,3]。

6.2　外显子改组的分子机制

关于外显子重组机制的研究最近有两个领域的工作值得关注:非常规重组和 LINE-1(L1, long interspersed nuclear element)单元调节重组。所谓非常规重组是很少或无顺序同源性之间 DNA 重组。

Bloemendal 等[5]揭示了非常规重组导致仓鼠热休克蛋白 α A-晶体蛋白基因的外显子改组。将仓鼠 α A-晶体蛋白基因转染到鼠肌肉细胞系中,鉴定了大量基因内复制的突变体[图 6.1(a)],这些突变体的顺序分析指出,它是通过两个 α A-晶体蛋白基因中两个 CCCAT 位点之间的非常规重组产生的,一个在内含子 3,而另一个在外显子 2(α A-晶体蛋白的另一种异构形式的内含子),结果形成内部复制的外显子结构。已经证实 3 个一组顺序 CAT 是优先被非常规重组中拓扑异构酶 1 打开缺口,包含非常规重组中两个 CCCAT。这样一种特殊的顺序要求,可能暗示低速率重组,它不能解释观察到的较高频率外显子改组。

另一种机制也能导致外显子改组,称为 L1 单元调节 3′转导[图 6.1(b)]。L1 是一种逆转录转座子(retrotransposon),能逆转录并在哺乳动物基因组中运动。Moran 等[6]在转染实验中证实 L1 由于其结合,从而弱化了终止信号。L1 可能以适当的频率产生通读转录。L1 单元在哺乳动物基因组中的高丰度(人基因组约 15%),L1 单元调节 3′转导似乎代表改组基因组顺序经常发生的一种机制。基因组分析揭示通过这种机制转导的顺序约占人基因组的 1/4。L1 含有 5′非翻译区(UTR),两个非重叠的可读框(ORF1 和 ORF2),和一个在 poly(A)尾端的 3′UTP。ORF1 编码一个结合 RNA 的蛋白质,而 ORF2 编码内切核酸酶(EN),逆转录酶

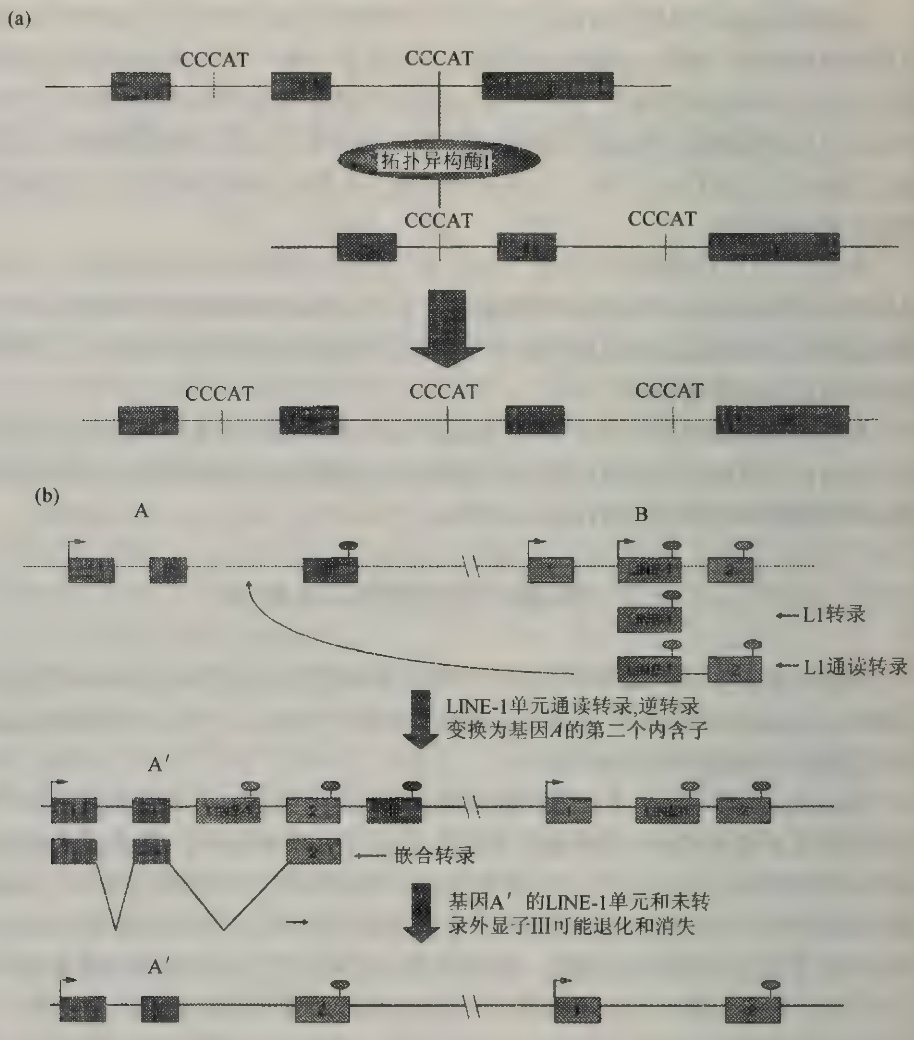


图 6.1 外显子改组的分子机制

(a)非常规重组导致超 α A-晶体蛋白的外显子复制。重组发生在超 α A-晶体蛋白基因内含子 1 和 11 中同源顺序 CCCAT 的两个非同源位点之间。拓扑异构酶 1 优选缺口顺序 CAT,形成两个新功能 CCCAT。(b)LINE-1 调节的 3′转导的外显子改组。为了说明转导过程,A 和 B 是两个假设的基因。A′是通过该过程产生的新基因。转录起始位点和终止信号分别以箭头和椭圆形表示。LINE-1 转录终止信号比较弱(椭圆形)产生通读转录(包括基因 B 的外显子 2)

(RT)和未知功能的富半胱氨酸(C)结构域。L1 逆录转座子被认为是移动的 Alu 和加工的伪基因(pseudogene)。

L1 可以逆转录转座于基因中;L1 可以容易地由其 3′侧把 DNA 转导到新基因

组位置;L1 调节转导可以产生新基因。产生新基因的机制包括点突变、外显子改组、DNA 复制、不对称交换、DNA 易位、定位重组和功能加工伪基因形成。L1 调节转导是一种基因组多样化的潜在来源(图 6.1)。它通过 RNA 中间体产生,不需要同源重组,并且允许非 L1 顺序分散到新基因位点。被改组的顺序可能是启动子,增强子或外显子,并且它们的分散可能导致新基因的产生或改变已存在的基因的表达。由于 L1 逆录移位往往是 5′-剪切,某些传导可以缺乏 L1 顺序。

外显子改组如何产生新基因? 为此 Patthy[7]研究了通过外显改组所产生的上百个蛋白质族(主要是哺乳动物基因编码的)内含子相(intron phase)(相被定义为内含子或在密码子之间或密码子内的相对位置)统计分析揭示出优势相 0。外显子和内含子相的关系,即外显子倾向定位于同一相内含子的两侧,这种内含子相的非随机分布指出真核生物基因的大部分源于外显子改组,这就是说外显子改组在新基因进化中起重要作用。

真核生物的进化是由亲代基因组有性重组调节的。交换体随机出现频率取决于 DNA 长度,由于外显子只占人基因的 1%,内含子约占 24%。这就是说内含子平均长度远大于外显子的平均长度,并且重组频率正比于 DNA 长度,所以大多数交换体发生在非编码顺序。此外,内含子含有大量的易位单元和重复顺序(例如 Alu 重复——能够被特异分裂 DNA 的酶切割称为 Alu),它促进非同源基因的错配和重组,这提示内含子是遗传重组的热点,外显子改组已成为蛋白质进化的主要因素。

6.3　蛋白质结构域是运动的功能模块[8]

结构域是蛋白质亚基结构中的紧密球状区域,是介于蛋白质二级与三级结构之间的结构层次。结构域在蛋白质中呈独立的结构单元,独立的功能单元和独立的折叠单元。在复杂蛋白质中,结构域具有结构模块(module)和遗传单元的作用。结构域又可分为亚结构域。

结构域广布于球状蛋白中,域间连接或松散或紧密。典型的双结构域蛋白质的域间连接呈铰链(hinge)式。结合底物或辅酶后伴随着域间的相对运动。在蛋白质中,不同的结构域具有特定的功能,例如结合底物,催化反应,亚基间相互作用及活性调节等。多结构域的活性部位大都是位于两个或多个结构域的界面之间。在催化作用进行的同时,必然伴随着这些结构域的相对运动,结构域间的松散连接为域间运动提供了可能性,以调节刚性(对稳定性而言)和柔性(对活性而言)之间的最佳平衡,通过刚性和柔性之间的表观交替变位提高酶对环境的适应性。在复杂蛋白质中结构域往往以结构模块方式组合。

生物进化过程的本质是分子进化。蛋白质的进化,特别是结构域的进化起核

心作用。通过不同结构域的组合可以形成功能各异的蛋白质分子。

蛋白质顺序和三维结构分析已经提示许多蛋白质是由一些分离的结构域组成的。这些所谓的嵌合蛋白质在后生动物中是特别丰富的。嵌合蛋白质的个别结构域往往涉及特殊功能。嵌合蛋白质中所发现的结构域在进化上是运动的,这意味着它们在进化过程中是分散的,现在出现在各种不相关的蛋白质中。运动的结构域是独立的折叠单元,主要是由于它们插入到新的蛋白质环境中防止错误折叠。运动的结构域示如表 6.1。

表 6.1　运动结构域与外显子结构的关系

结构域	外显子分类	蛋白质
类 EGF 结构域	1-1	表皮生长因子前体
	1-1	TPA
	1-1	因子 X
环饼结构域	1-2-1	肝细胞生长因子
	1-2-1	TPA
	1-2-1,1-1	凝血酶原
纤维结合素 I 型结构域 (指状结构域)	1-1	纤维结合素
	1-1	因子 XII
	1-1	TPA
纤维结合素 II 型结构域	1-1	纤维结合素
	1-1	巨噬细胞露糖受体
	1-2-1	因子 XII
纤维结合素 III 型结构域	1-1-1;1-1;1-0-1	纤维结合素
	1-1-1;1-0-1	axonin
	1-0-1;1-1-0-1	Kallmann syndreme 蛋白质
PAN 结构域	1-2-1	血纤维溶解酶原
	1-2-1	Prekallikrein
	1-2-1	因子 XI
Gla 结构域	1-0-1	因子 X
	1-0-1	凝血酶原
	1-0-1	蛋白质 S
Sushi 结构域 (SCR 重复, CCP 模块)	1-1	因子 XIIIb
	1-1	P 选择素
	1-1-1	肝球蛋白

审查编码嵌合蛋白的基因揭示出结构域与外显子结构之间的基本关系，即每个结构域倾向于由同一拼接框中开始和终止的一个或多个外显子组合编码。也就是说是嵌合蛋白质似乎是通过外显子改组由组合在一起的多个结构域产生的。大多数内含子是“假 DNA”(junk DNA)，高度耐受缺失和插入，中间重复顺序(例如 Alu 重复)可以插入内含子的上述区域，这可以增加错误匹配和非同源内含子重组的机会。前 mRNA 内含子的重要特征：不同内含子的拼接点是等价的，之间可交换，这对外显子改组是重要的。模块蛋白质的进化分析表明外显子改组主要在动物界。

运动的模块与蛋白质的结构域的区别是它的运动性；模块是位于同一相内含子的两侧以促进它的分散。根据这一定义，如果同一相的内含子被插入到它的两个边界上的话，任何蛋白质结构域都可以转为模块。图 6.2 表示结构域转为运动的模块的过程。

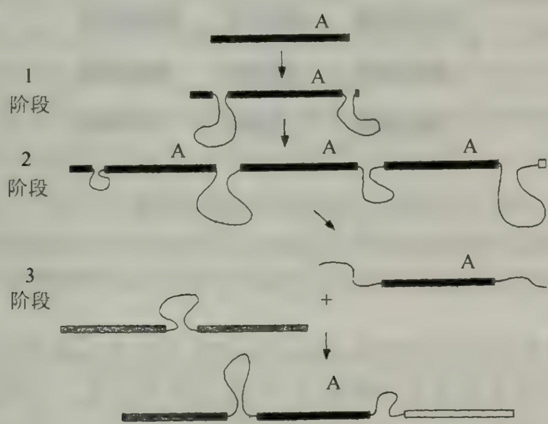


图 6.2　结构域转为运动的模块的不同阶段

图中指出具有分泌信号肽结构域的蛋白质的模块化。阶段 1：在结构域的 N 端和 C 端边界上插入同一相内含子；阶段 2：对称(0-0,1-1 和 2-2)的原始模块内部串联复制；阶段 3：在含有一些串联拷贝(例如多聚模块蛋白质)的基因压缩时，模块往往被切割，这样提供运动的模块，被切割下来的模块可以插入到其他基因中

3 个对称的模块组(1-1,0-0 和 2-2)对外显子改组来讲是相等的，因为它们同样满足内含子重组的相适应性规则，然而 1-1 模块却存在优势。

这种模块化假说可分为三个不同的阶段：第一阶段，内含子插入相应蛋白质结构域的基因片段边界上，内含子可以作为可动的遗传单元并转移到其他的基因组位点；第二阶段，插入的内含子重组，串联复制形成所谓“原始模块”(protomodule)；第三阶段，一个或更多的原始模块通过内含子重组转移到不同的非同源基因上。

外显子改组是受拼接框规则控制的。内含子可以插入在两个连续的密码之间

基因阅读框(定义为相 0 内含子)或一个密码子的第一和第二核苷酸之间(定义为相 1 内含子)或第二和第三核苷酸之间的基因阅读框(定义为相 2 内含子)。根据内含子两侧的相,外显子可以分为 9 个不同的组(图 6.3)。这些包括对称的外显子 0-0,1-1 和 2-2;非对称外显子 0-1,0-2,1-0,1-2,2-0 和 2-1,它可以连成对称的组合。不对称外显子(或外显子组)可以插入同一相内含子,经历串联复制或缺失干扰阅读框。

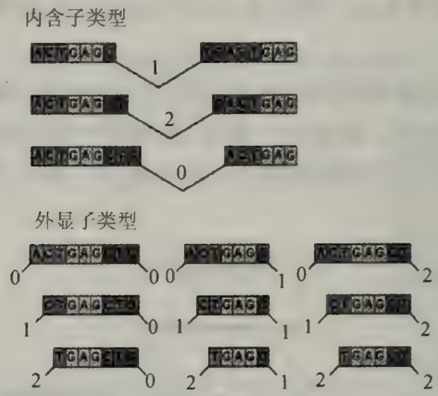


图 6.3　内含子和外显子类型

内含子可以分为相 0,相 1 和相 2 内含子,这取决于它们相对于基因可读框的位置。外显子可以分为 9 种不同的类型,这取决于它们位于内含子两侧的相

表 6.1 中大多数蛋白质结构域是由对称外显子或非对称外显子的对称组编码的。由于这些结构域自身含有功能模块,可以在周围运动。它们在蛋白质进化中起着支配作用。值得强调的是嵌合基因的原始内含子-外显子组织可能通过丢失或获得内含子或通过拼接连续滑动,在进化过程中模糊不清。在某些编码结构域的外显子边界上缺乏同一相内含子可能是这些外显子已经历外显子改组。目前已经鉴定出 60 多种运动结构域是由 1-1 外显子(或外显子组)编码,而 0-0 和 2-2 是少数。1-1 结构域占优势的原因是不知道的,可能与增加产生嵌合蛋白效率有关。大多数嵌合蛋白是细胞外或构建膜蛋白的胞外部分,因此,已经指出嵌合蛋白在多细胞进化中起重要作用。

6.4　外显子改组产生模块酶[8,9]

6.4.1　血液凝集和血纤维溶解的蛋白酶

血液凝集和血纤维溶解的完整过程形成级联酶促反应,其中无活性的酶原转

为活性酶。这些酶原属于丝氨酸蛋白酶族，它们的活化伴随着有限量肽链的水解。血液凝集和血纤维溶解蛋白酶(BCFE)的氨基酸顺序同原型丝氨蛋白酶例如胰蛋白酶和胰凝乳蛋白酶比较表明，这些酶有延长的 N 端，其由一些功能确定的分离结构域组成(图 6.4)，例如底物识别，与磷脂膜的结合，或同蛋白质辅因子相互作用。N 端结构域包括类 EGF 结构域、Gla 结构域、纤维结合素 I 型和 II 型结构域、环饼结构域和苹果状结构域。这些结构域也存在不相关的蛋白质中，像脱辅基脂蛋白 A 中存在环饼结构域，纤维结合素 I 型和 II 型结构域。

BCFE 的结构域组织表明了结构域与外显子结构是密切相关的(图 6.3)。类 EGF 结构域完全是由单一外显子编码，其属于 1-1，并且内含子—外显子接合点定位在结构域边界的 3 个氨基酸(C 端或 N 端)之内。纤维结合素 I 型结构域(也称指状结构域)也是由单一的 1-1 外显子编码。5′外显子编码前肽部分以使其分泌到细胞质中。BCFE 中大多数环饼状结构是由两个外显子的对称组编码(1-2-1)。然而，在凝血酶原中，第二个环饼结构域是由单一的 1-1 外显子编码，显而易见，插进的相 2 内含子在进化过程中丧失。编码血纤维蛋白溶酶原的环饼结构域的两个外显子是由相 0 内含子分开。这种相移位可能是拼接点滑动的结果。纤维结合素 II 型结构域和苹果形结构域是完全由两个外显子对称组(1-2-1)编码。研究表明苹果形结构域同血纤维蛋白溶酶原的前体肽(PAP)和少数其他的结构域有明显的同源性，这个结构域族称为 PAN 模块族。上述资料指出编码 N 端结构域所有的外显子组是位于相 1 内含子的两侧。这个观察对外显子改组产生 BCFE 提供额外的证据。

除 BCFE 外，几个其他的细胞质蛋白质也含有运动的结构域。例如，蛋白质辅因子的因子 V 和 VIII，它们的 C 端区都含有两个 C 结构域。这个结构域也在其他非同源蛋白质(neuropilin 和 discoidin)结构域受体中发现，是由 4 个外显子对称组(1-0-0-1)编码，约 60 个氨基酸的 10 个串联重复，称为 Sushi(SCR)结构域，已经在转谷酰胺酶因子 XIII 的 β 链中鉴定出。每个结构域是由单一 1-1 外显子编码。肝球蛋白和少数其他蛋白质也含有 Sushi 结构域，它是由两个外显子对称组编码。Von willebrand 因子的前体是由 4 个不同的重复结构域(A，B，C 和 D)组成，每个以 2～5 个拷贝存在。结构域 A 和 C 也出现在各种其他的嵌合蛋白质中，包括补体因子、整合素、胶原蛋白和凝血因子(thrombospondin)。在第八因子相关抗原(VWF 因子)中，A 结构域的头两个重复是由单一 1-1 外显子一起编码，而第 3 个重复是由 4 个外显子(1-1-1-1)编码。关于 C 结构域，结构域组织与外显子结构之间没有明确的关系。Von willebrand 因子 C 结构域也在 nov 蛋白质和胶原蛋白 α1(II)链中发现，然而，它是由 1-1 一个外显子编码(表 6.1)。通过外显子改组组装 BCFE 过程示如图 6.5。

模块组装假说的第一个证据是 tPA 的一个区与纤维结合素的指状结构域同

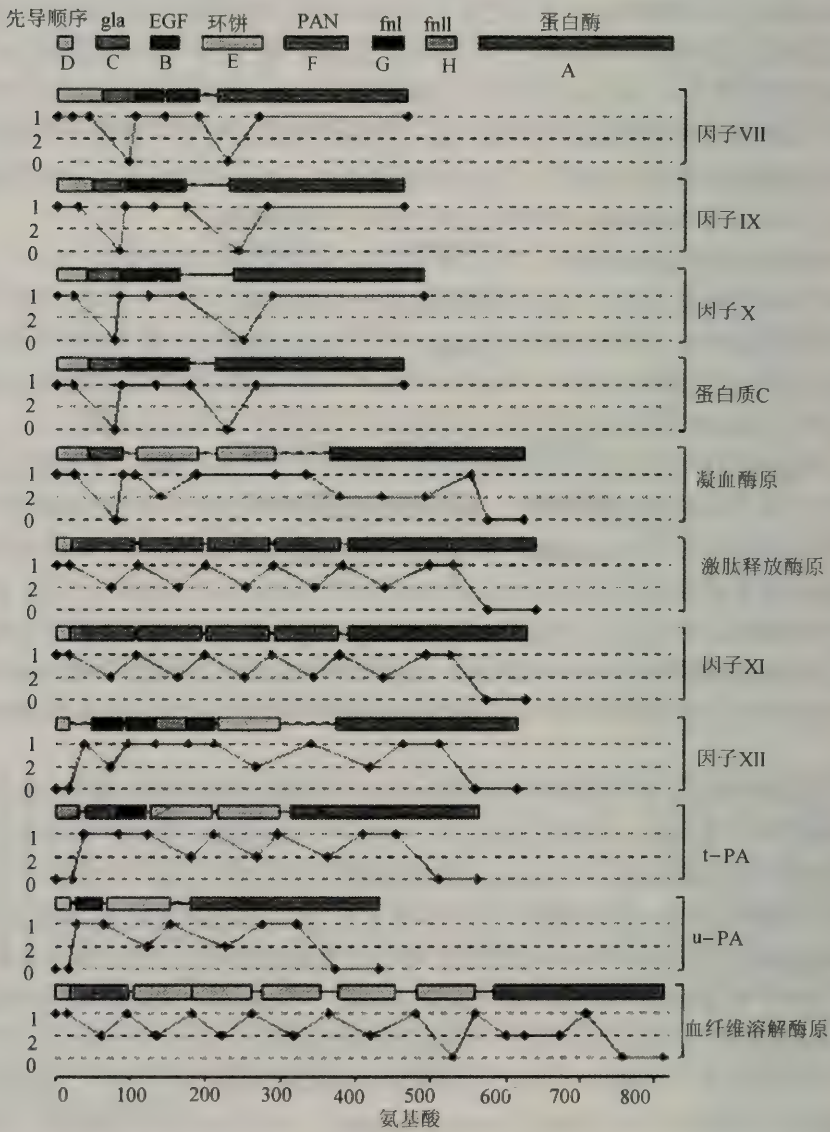


图 6.4　血液凝集和血纤维溶解的调节蛋白酶的结构域结构和拼接框图解

英文大写字母表示各种结构域：A，丝氨酸蛋白酶结构域；B，类 EGF 结构域；C，Gla 结构域；D，前-原先导顺序；E，环饼结构域；F，PAN 结构域；G，纤维结合素 II 型结构域（fnII）；H，纤维结合素 I 型结构域（fn I）。内含子-外显子接合点的位置以菱形实线表示，插入的内含子相以垂直轴表示

源;tPA的另一个区与表皮生长因子同源。这样很可能调节血液凝集和血纤维蛋白质水解的蛋白酶是外源模块组装的。调节蛋白酶的共同祖先是与简单的类胰蛋白酶相似,所有的调节模块是指被插入在信号肽和祖先蛋白酶原活化结构域之间的结构。组装过程涉及各种模块插入、串联复制和缺失。模块组装意味着编码模块的基因片段在进化过程中是运动性的(图6.5)。它们经历重复插入、串联复制、交换和缺失。外显子改组可以认为是这些结构域独特运动性的机制。内含子应当位于这些运动模块的边界上。tPA的嵌合结构及其基因的外显子-内含子组织之间关系提供了通过外显子改组进行模块交换的明确依据。

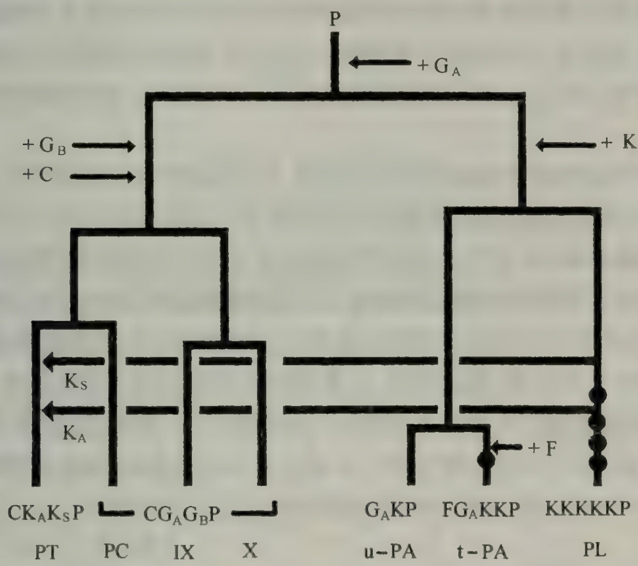


图6.5　由模块组装血液凝集和血纤维蛋白水解的蛋白酶的方案

图中缩写, P为胰蛋白酶同源的丝氨酸蛋白酶的祖先; GA,GB为A型和B型表皮生长因子模块; C为依赖维生素K的钙结合模块; K为环饼模块; F为指状模块; PT为凝血酶原; PC为蛋白质C; IX,X为因子IX,因子X; u-PA, t-PA为分别表示尿激酶、组织型血纤维蛋白溶酶原激活剂; PL为血纤维蛋白溶酶原。箭头指示模块插入, 实圆表示内部模块复制[10]

运动的模块可以为3种类型(类型1-1、类型0-0和类型2-2),在嵌含蛋白质进化中所用的大多数模块属于1-1组,2-2组或0-0组相对少。由外显子改组所产生的模块蛋白质主要是胞外蛋白质(体液蛋白质),胞外基质构建和转膜蛋白质的胞外部分,后者适应于胞外空间的氧化环境。许多模块富含二硫键(例如环饼、补体B模块、表皮生长因子模块、指状模块、纤维结合素II型模块等)。然而,通过外显子改组组装的所有蛋白质的一个共同的特征局限于动物,原核生物和植物或真菌则没有。纵观模块蛋白质,模块类型都表现出这种趋势:这些模块具有运动性和串联复制。

6.4.2 模块交换的原理

模块蛋白质(凝血酶原、因子 IX、因子 X、蛋白 C、血纤维蛋白溶酶原、尿激酶、tPA、白细胞介素 α-受体、肝球蛋白、纤维结合素及 LDL-受体)的基因结构揭示内含子参与组装的某些特征。内含子定位在环饼、生长因子、指状、纤维结合素 II 型、III 型,补体 B 和 LDL-受体模块的 5′和 3′边界,这些蛋白质总是拼接在相 1 可读框,即密码子的第一和第二个核苷酸之间,对相 1 内含子的偏好是相当引人注目。因为在非模块蛋白质基因中几乎不存在对相 1,相 0 和相 2 内含子偏爱问题。必须指出相的分类是内含子相当保守的特性。由于这种保守的偏爱,改变相的分类是不大可能的事情:内含子的两个拼接接合点经历同时发生和补偿改变,避免破坏蛋白质可读框。

相 1 内含子是在各种模块的 5′和 3′边界上。模块改组(通过内含子重组发生模块的插入,缺失和复制)的基本步骤是一个内含子的 5′拼接点与另一个内含子的 3′拼接点连接。显而易见,内含子拼接点的相容性对于内含子重组的进化价值是主要的:属于不同相的内含子重组将移动可读框,并将导致重组点下游的蛋白质信息的损失。换句话说,它是外显子改组绝对必要的条件,当相同相内含子重组拼接时,连接点一定连接。通过内含子重组、模块的插入、缺失和复制可以满足这个标准。模块在方向上是对称的,模块两侧的两个内含子是同一相(图 6.6 上部)。在原理上,这些对称的模块组是 1-1,0-0 和 2-2,这取决于相 1,0 或 2 内含子是否在模块的两侧。非对称模块(0-1,0-2,1-2,1-0 等)在实践上对模块蛋白质进化是无价值的。

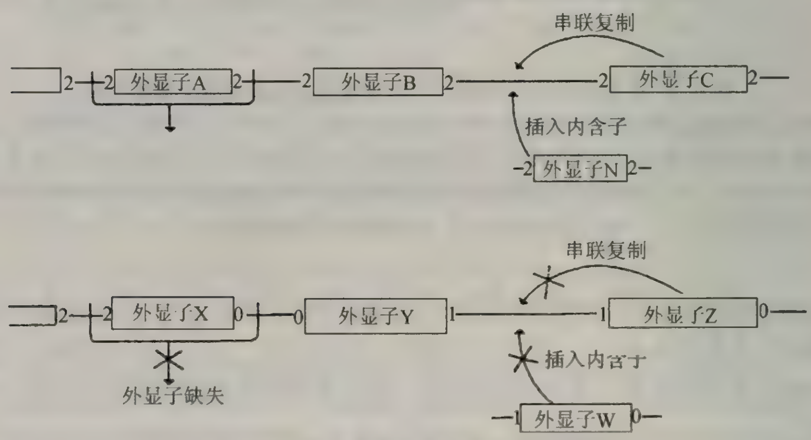


图 6.6 外显子改组中内含子相的意义

盒表示外显子(外显子组),连线表示内含子,数字表示相。如果模块是对称的(即当内含子在模块的 5′和 3′边界上是同一相)并且接受的也属于同一相,那么可读框不受影响(上部)。然而非对称外显子的缺失,复制或插入(即内含子在外显子 5′和 3′边界上是属于不同相)将破坏可读框(下部)[10]

1-1 模块:

根据上述分类,环饼、生长因子、指状结构模块,补体 B 模块等用于血液凝集,血纤维蛋白水解和补体活化的蛋白酶的构建,都是属于 1-1 组。根据相适应性规则,这些模块插入到祖先蛋白酶基因中需要相 1 靶内含子。这个靶内含子定位于祖先共同蛋白酶的信号肽和催化结构域之间。相 1 内含子是在丝氨酸蛋白酶族所有成员的胰蛋白酶同源区的 N 端边界上,这些包括简单的蛋白酶,例如弹性蛋白酶、胰凝乳蛋白酶、胰蛋白酶、激肽释放酶等。这个相 1 内含子是在它们的信号肽和胰蛋白酶同源结构域之间,这是祖先蛋白酶的所有"后代"的共同的内含子。胰蛋白酶同源区可能由 6 个内含子组成,内含子插入和去除结果形成目前的各种蛋白酶的外显子-内含子结

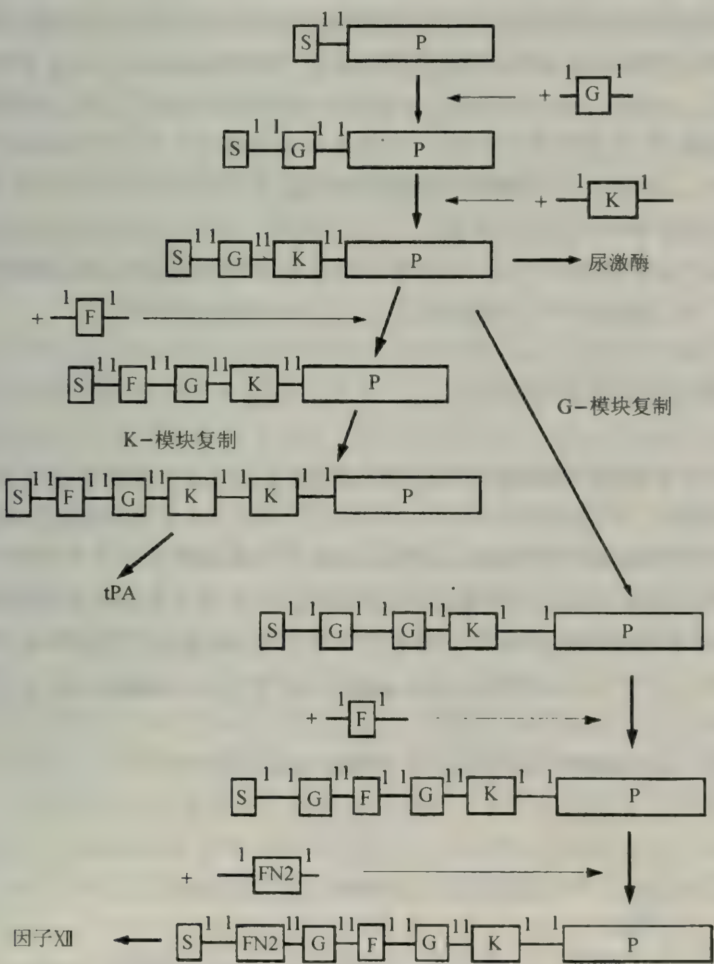


图 6.7　尿激酶、tPA 和因子 XII 通过 1-1 模块组装过程
数字表示参与组装过程内含子相。S—信号肽结构域;P—胰蛋白酶同源模块;
G—生长因子;K—环饼模块;F—指状模块;FN2—纤维结合素 II 型模块

构。祖先蛋白酶在其信号肽和蛋白酶同源结构域之间有相 1 内含子。1-1 相内含子总是插到这里,这个相 1 靶内含子易接受进来的 1-1。

以尿激酶, tPA 和因子 XII 作为例子, 它们详细组装过程示如图 6.7, 由图可见, 1-1 生长因子,环饼、指状和纤维结合素 II 型模块的插入和复制被分开,最终导致基因结构中所有模块间的内含子属于相 1 类。相类优势是外显子改组进化的必然结果。外显子改组不仅反映了蛋白质结构域组织及其基因的外显子-内含子组织之间的关系,而且也反映了模块间内含子的特征相偏向。模块蛋白质基因的结构遵循这些规则,实际上是外显子改组进化的证据。这个规则的正确性是由几个蛋白质的嵌合性质所证明。

相适应性规则的必然结果是 1-1 模块只能与其他 1-1 模块连接,而不能与 0-0 或 2-2 模块连接。这样可以由 1-1 模块构建模块蛋白质簇(cluster)。在这个簇内模块的交换只能是 1-1 模块,而不是 2-2 或 0-0 模块。选择素族提供了这样一种关系的典型例子:C 型凝集素、生长因子和补体 B 型模块是由分离 1-1 内含子编码的,即模块间内含子是相 1。然而也有少数情况,基因的结构与蛋白质的模块组织不是这种关系。

通过内含子插入可以改变外显子-内含子结构。内含子去除在改变基因外显子-内含子结构中也起重要的作用。逆录移位是科学家所熟悉的机制,基因的内含子可以被消除。与大量的各种 1-1 模块相反,2-2 或 0-0 模块相对较少。

6.4.3　模块交换的机制

通过内含子重组使模块缺失和复制的详节见图 6.8。正常和非正常 LDL-受体基因结构比较揭示了模块两侧的内含子中间重复顺序(Alu 重复)的功能,错误匹配和重组负责个别模块的缺失或复制。例如,LDL-受体基因内含子 1 和 3 中 Alu-Alu 重组是负责第二和第三 LDL-受体模块的缺失;内含子 4 和 5 中 Alu-Alu 重组负责缺失编码第 6LDL-受体模块的外显子; 内含子 1 和 8 中 Alu 重复负责复制 9 个模块(7 个

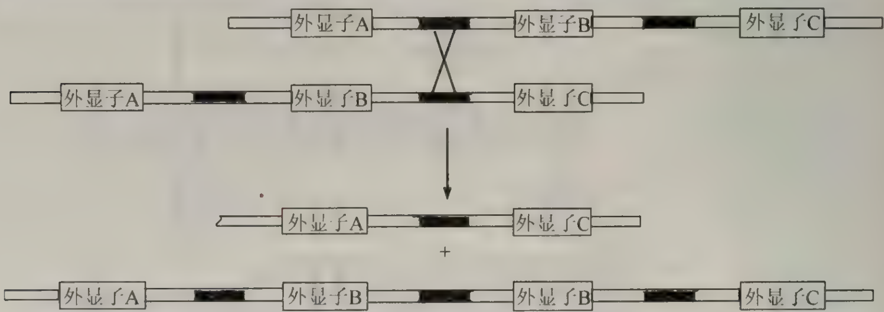


图 6.8　通过内含子重组产生模块的复制和缺失

内含子内的黑盒代表中间重复顺序(例如 Alu 重复)

LDL-受体模块和两个表皮生长因子模块)。内含子 6 和 14 的 Alu-Alu 重组负责 LDL-受体完整的表皮生长因子前体同源结构域的缺失。

模块缺失或复制的机制与影响 LDL-受体基因的非模块部分 Alu-调节内含子重组相同。两者的主要区别是:在模块蛋白质中,改变模块数产生具有活性的蛋白质,而在非模块蛋白质中,折叠结构域的缺失似乎破坏了蛋白质的完整性。

模块如何由一个基因运动到另一个基因? 一种似乎有道理的假设是与内含子重组机制相同,模块串联复制和缺失也负责它们的转移,然而也有其他的可能性。模块的插入可能与内含子插入机制相同。但必须强调,在“内含子重组模型”和“模块改组逆录移位模型”之间有重要区别。第一种情况的内含子必须遵守相适应性规则,而第二种情况没有这种要求。模块被插入到目标基因预先存在的内含子中,通过内含子重组而复制和缺失也使它们由供体转移到受体基因。对称模块往往经历重复串联复制,结果形成含有一些串联拷贝的模块蛋白质。

6.5　体外外显子改组

6.5.1　策略和方法

外显子改组的自然过程可以体外模拟,通过体外外显子改组库和筛选进化酶。体外外显子改组的策略和方法示如图 6.9。在这种方法中,应用嵌合寡核苷酸扩增,编码结构域的外显子或外显子组合,然后这些 PCR 片段的混合物应用自我引动重叠聚合酶反应组合组装成全长基因。当在一个基因的外显子与另一个基因的外显子连接时,发生重组。

外显子改组库的多样性是由一些被变换的模块数和每个模块所用的同源物数决定的。在嵌合寡核酸设计中,只选择连接的寡核苷酸。按不同摩尔比混合编码结构域的 DNA 片段,人们可以精确地控制最终库中个别结构域的化学计量,像在自然进化一样,可以通过编码结构域的外显子插入、缺失或顺序改变来增加多样性。

构建外显子改组库有几种方式(表 6.2)。高质量的库是通过改组编码结构域的外显子构建的。这样只获得少数的交换体并保持大部分连接信息。然而,另一种相反的方式通过完全变更外显子构建蛋白质库。这个途径破坏了所有的先前存在的连接信息,结果形成低平均功能库,但可以产生稀有的高新组合。构建高功能库优选密切相关种类同源物取代一个或更多编码结构域的外显子。例如,人的结构域可以用灵长目动物和其他哺乳动物同样蛋白质的同样结构域取代。因子 VIII 的报道表明相关种类的结构域取代可以改进蛋白质功能。已证实人的因子 VIII 的稳定性,并且当 Al 和 A2 结构域或只是 A2 结构域被猪的因子 VIII 相应结构域取代时它的凝集活性增加。对蛋白质药物而言,为了使免疫反应减少到最小,人们

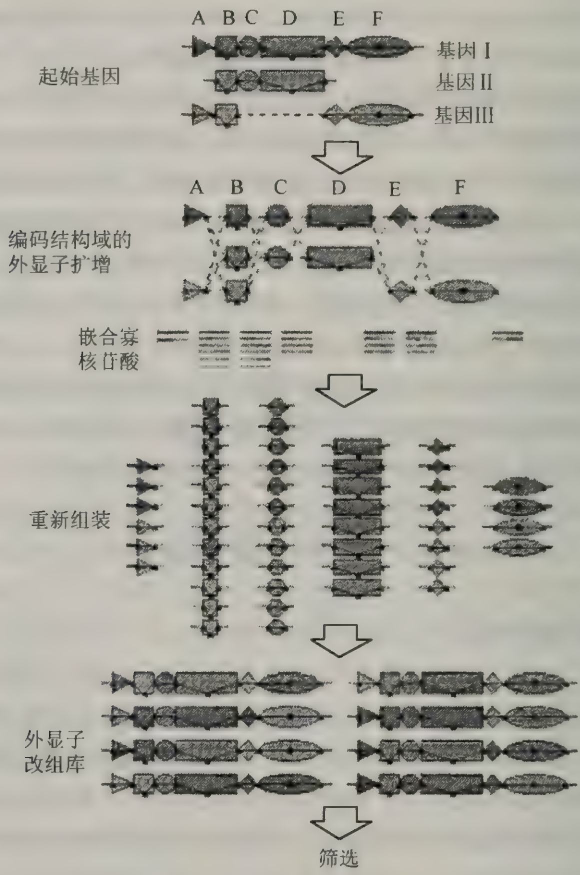


图 6.9　体外外显子改组

这种方式分 3 步：①编码结构域的目标基因的外显子（或外显子组）和相关基因的同源物应用嵌合寡核苷酸扩增；②1 中合成的 DNA 片段作为类 PCR 反应的引物和模板，重新组装成全基因；③根据要求筛选外显子改组库

表 6.2　外显子改组库的类型

不同种外显子改组	外显子用不同种的同样基因的等价外显子取代
同种外显子改组	外显子用同种的不同基因的同源外显子取代
不同种结构域改组	结构域用不同种的同样基因编码的等价结构域取代
同种结构域改组	结构域用同种不同基因编码的同源物取代
结构域数改变	蛋白质结构域的缺失，复制或插入
功能同源物改组	结构域用功能上相关的非同源结构域取代
从头蛋白质组装	多个独立的结构域组装成新的蛋白质

可以构建只含有人外显子的基因库。保守的途径是用来自不同人基因的同源外显子取代一种人的外显子或外显子组(即同种间取代库)。这个途径可以提高酶的活性或改进酶的性质。

外显子改组库可以由没有顺序或结构同源性的结构域交换产生。在这个途径中,改组所用的结构域可以根据它们的功能性质来选择。Langer-Safer 等[10]构建了一个 tPA 突变体,其中类 EGF 结构域和纤维结合素 1 型结构域是用血纤维蛋白溶酶原的非同源环饼 1 结构域取代。去除两个 N 端结构域反而增加了该酶在细胞质中的半衰期。对消化血纤维蛋白的血纤维蛋白溶酶的亲和性是由插入的血纤维蛋白溶酶的环饼 1 结构域保持。

功能克隆的分级将取决于改组过程的保守性。由各种保守水平、范围从同种(免疫同源性)的同源物取代外显子到不同种的同一基因的同源外显子,到顺序同源性和功能同源性产生简单的相容拼接框。

合理设计仅能构建少数的特殊的嵌合突变体,外显子改组可以产生各种酶的大库,由筛选鉴定最好的克隆,然后再经几轮突变以达到最佳化。

6.5.2　组合结构域重组

分子生物学方法和高通量筛选的进步以及自然界蛋白质进化中模块组装的灵活性和多样性,有力地促进了组合结构域重组各种技术的发展,其中最好的是体外外显子改组。

关于体外外显子改组,早期集注在催化 RNA 库融合的自身拼接内含子的探索方面。Fisch 等[11,12]构建两个随机 30 个碱基对外显子库,带有插入 Lox-Cre 重组位点的组 I 内含子顺序位于随机碱基对顺序的 5′端或 3′端。第一个库是携带外显子库的质粒转化到 *E. coli*,第二个库是通过噬菌体感染转化。体外通过 Cre 重组酶重组两个库产生约 1.6×10^{11} 库。在两个外显子库之间的 Lox-Cre 重组被相邻内含子顺序切除,在融合点上保留 15 个碱基对。筛选肽库,用噬菌体展示鉴定出几种顺序,对各种靶蛋白质结合力达 nmol/L。同时,Mikheeva 和 Jarrel[13]提出直接通过工程反式拼接组 II 内含子核酶进行体外外显子改组。这个方法是在人 tPA 两个片段组成的模型系统进行的。RNA 类似物是通过两个外显子分开逆转录产生的,与它们相应的内含子顺序融合。两个转录产物混合物反式拼接。虽然连接效率是低的,但凝胶分析明确地指出产物的形成。与以前的途径相反,工程核酶促进了两个外显子融合。

内含子催化重组已由 DNA 改组衍生出新方法[14,15]。最初实验是用人 *sFv* 进行的。在该结构中有 6 个互补决定区(CDR)。DNase I 消化母体基因产生的片段用合成的寡核苷酸桥连,混合物用 PCR 扩增,库分析指出野生型和突变体 CDR 的

随机混合分布在所有 6 个区。这种方法为嵌合寡核苷链节设计库给出很大的灵活性。链节的大小和顺序以及每个库中外显子的浓度，不同库的混合比例都可以控制组装过程(图 6.10)。



图 6.10　外显子改组

多结构域酶的组合库可以通过多个母体(A～D)的模块或结构域重组以有序的方式产生，保持母体大小(左图)或随机组装，允许在单一步骤中模块的插入，缺失，取代或复制。图中黑、白六角形，星和圆表示互补单元(互补 DNA 股或反式拼接蛋白质内含子)

通过反式拼接蛋白质内含子进行后翻译蛋白质片段的重组可以用于组合蛋白质模块[13]。在第一个蛋白质片段的 C 端和第二个蛋白质片段的 N 端的短蛋白质顺序(蛋白质内含子)翻译、二聚合并催化两个蛋白质片段之间肽键的形成而同时自身切割(图 6.11)。顺式拼接蛋白质内含子的应用也有报道[14]，Otomo 等[16]报道

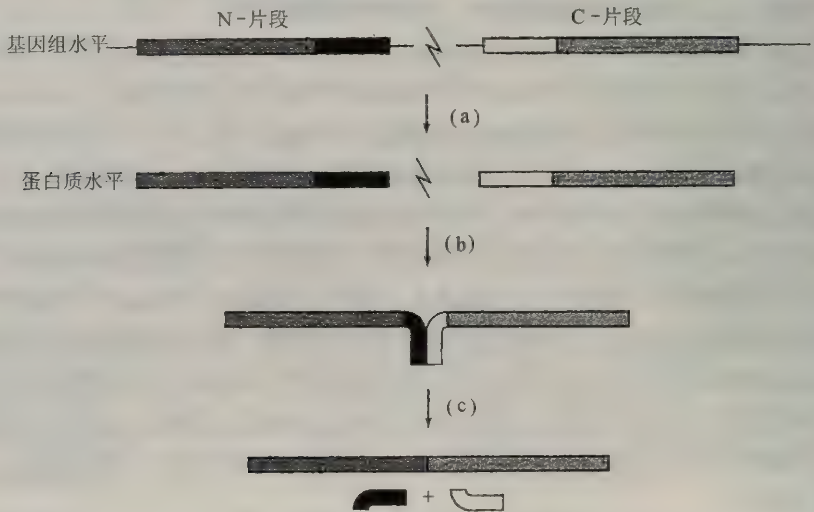


图 6.11　通过反式拼接蛋白质内含子连接两个蛋白质片段

(a)在 DNA 水平，两个片段可以被分开的可读框编码，位于 3′延伸编码的 N-蛋白质内含子(N-片段)或 5′延伸编码的 C-蛋白质内含子(C-片段)；(b)转录和翻译，蛋白质内含子结构域二聚合；(c)自身切割，N-片段和 C-片段融合

了用两个分开的反式拼接蛋白质内含子对标签化的3个蛋白质片段的体外重组、连接、体外再折叠，正确地组装成全长蛋白质。

6.5.3 体外外显子改组实例

实例1. 超α A-晶体蛋白，体外外显子改组的产物[5]

已知当质粒DNA通过哺乳动物细胞时，转染的DNA中所有突变都可以自发地发生。因此，当基因转染到真核细胞中产生畸变(aberrant)产物，按逻辑应是不理或弃掉。然而，转染和其他的重组DNA技术可以简单地开发出正常发生在生命有机体的突变过程，可以增加我们对突变和进化机制的了解。

为了比较细胞内各种哺乳动物小的热休克蛋白质(sHsps)的性质，按常规将相应的基因转染到不同细胞系中。在6种已知的哺乳动物sHsps中，一种特殊的位置是由α-晶体蛋白质占据。α-晶体蛋白质是第一个作为主要结构的眼球晶状蛋白质被发现的，它是以大的球状复合物(800kDa)存在，含有两种类型的亚基(20kDa)，α A-晶体蛋白质和β B-晶体蛋白质。在啮齿动物和某些其他哺乳动物中，存在较小的α-晶体蛋白质亚基，起自α A-晶体蛋白质的基因转录的选择拼接。这个过程除了产生正常的173个残基α A-晶体蛋白质外，还有具有23个氨基酸残基插入的延伸α A^{ins}晶体蛋白质。这个插入是由外显子2编码，由于相邻非一致GC5′拼接位点的存在，它在80%～90% mRNA中被跳读。像sHsps族的其他成员，α A-晶体蛋白质在体外有类分子伴侣活性，抑制变性蛋白质的凝集。α-晶体蛋白质在体内最明显的特征是它们能够在不同的细胞类型超表达的情况下赋予耐热性。

α B-晶体蛋白质除了晶状体外在许多组织中存在，在心脏和条纹肌更多，α A-晶体蛋白质主要局限于晶状体。为了测定是否α A-晶体蛋白质进化过程中在功能上是α B-晶体蛋白质趋异进化。有趣的是α A-晶体蛋白质在各种细胞类型中表现是α B-晶体蛋白质的正常结构域。在肌肉细胞系(一个稳定的转染细胞系)表达，除了预期的α A-晶体蛋白质和α A^{ins}-晶体蛋白质外，还高水平的表达了两个免疫学上交叉反应的蛋白质。实验发现这两个稍大的蛋白质是由大的基因内复制的突变体α A-晶体蛋白质基因编码的，在两个CCCAT同源物非常规重组产生的(见图6.1)。串联复制的外显子3顺序存在这个基因成熟mRNA中，结果在翻译蛋白质中形成41个残基的重复。细胞表达的延伸α A-晶体蛋白质有正常的生长特性，并且通常扩散到细胞质。突变体蛋白质容易移入正常的大的水溶性α A-晶体蛋白质复合物，这表明插入不干扰复合物的整合。这样，α A-晶体蛋白质突变体模拟外显子复制是哺乳动物基因组进化中外显子改组的结果。

实例2. 外显子交换产生新功能蛋白质[17]

真核生物基因外显子和内含子结构的发现进一步阐明了基因的进化。外显子

是编码蛋白质分子的功能单位，并且新功能蛋白质是由各种内含子交换体产生的功能单位组合选择所得到的，即是由一些外显子模块进化来的。为了执行生物功能，蛋白质必须具一种稳定的特殊构象。如果外显子编码结构和功能单位，那么，外显子的组合可以预期能够产生稳定的功能蛋白质。“模块”结构作为一种紧密的蛋白质结构单元，由残基的 α-碳原子之间的距离鉴定之，并与基因的外显子区相适应。基因外显子区域和模块蛋白质结构之间的紧密关系为 Black 观点提供了支持[2]。

α-乳清蛋白是一种哺乳动物乳清蛋白质，并是乳糖合成酶(EC2.4.1.22)的一个调节亚基，当其与半乳糖苷转移酶结合时，它催化半乳糖转为 *N*-乙酰-D-氨基葡萄糖(CLcNAc)，催化 UDP-半乳糖和葡萄糖有效地合成半乳糖。α-乳清蛋白也与钙强烈地结合。α-乳清蛋白的一级和三级结构与 C-型溶菌酶同源，但功能上有很大的差别。鸡卵清溶菌酶(EC3.2.1.17)催化多糖中 GLcNAc 和 *N*-乙酰-D-胞壁酸(MurNAc 之间的 β-1,4 糖苷键水解。它的结构与功能关系及基因组织已阐明。溶菌酶的催化部位定位在基因的外显子 2 上(编码残基是 Trp2—Ala82)，其包括催化活性残基和结合寡糖底物环 C、D、E 和 F 的一簇残基。鼠、几内亚猪、人和山羊 α-乳清蛋白基因与鸡卵清溶菌酶基因组织紧密相关。α-乳清蛋白质和 C-型溶菌酶基因在精确的同样位置上有 3 个内含子，但内含子的结构和 4 个外显子稍有区别(图 6.9)。因此，两者非常相似，它们是由共同祖先通过基因复制趋异进化而来的，它们是蛋白质结构与功能关系进化研究的工具。

鸡溶菌酶外显子 2 区的大多数氨基酸残基组成它的催化裂缝。羊 α-乳清蛋白质的外显子区用鸡溶菌酶的外显子区取代，赋予 α-乳清蛋白催化活性(α-乳清蛋白在其天然结构形式上是一个非酶促蛋白质)。这说明通过外显交换可以产生新功能的蛋白质。

目前，关于外显子改组，体内研究报道较多，体外研究报道较少，是一刚刚兴起的研究领域。蛋白质数据库研究表明外显子改组是新基因起源的普遍机制。因此体外模拟体内外显子改组将是定向进化普遍应用的方法[18~22]。

参 考 文 献

1 Gilbert W. Why genes in pieces. Nature, 1978, 271: 501

2 Blake C C. Exons encode protein functional units. Nature, 1979, 277: 598

3 Roy S W, Nasaka M, de Souza S J, *et al*. Centripetal modules and ancient introns. Gene, 1999, 238: 85~91

4 Long M. Evolution of novel genes. Current opinion in gentic development, 2001, 11: 673~680

5 Van Rijk A F, Van den Hurk M J, Renkema W, *et al*. characterists of super α A-Crystallin, a product of *in vitro* exon shuffling. FEBS Lett, 2000, 480: 79~83

6 Moran J V, De Berardinis R J, Kazazian Jr H H. Exon shufflig by L1 retrotransposition. Science, 1999, 283: 1530~1534

7 Patthy L. Genome evolution and the evolution of exon-shuffling-a review. Gene, 1999, 238: 103 ~ 114

8 Kolkman J A, Stemmer W P. Directed evolution of proteins by exon shuffling. Nat Biotechnol, 2001, 19: 423 ~ 428

9 胡红雨,许根俊. 蛋白质的结构转换. 生物化学与生物物理进展,1999,26:9 ~ 12

10 Patthy L.(eds), Protein evolution by exon-shuffling. Springer-verlag, 1995, 1 ~ 135

11 Langer-Safer P R, Ahern T J, Angus L B, *et al*. Replacement of finger and growth factor domains of tissue plasminogen activator with plasminogen kringle. J Biol Chem, 1991, 266: 3715 ~ 3723

12 Fisch I, Kontermann R E, Finnern R, *et al*. A strategy of exon shuffling for making large peptide repertoires displayed on filamentous bacteriophage. Proc Natl Acad Sci USA, 1996, 93: 7761 ~ 7766

13 Mikheeva S, Jarrell K A. Use of engineered ribozymes to catalyze chimeric gene assembly. Proc Natl Acad Sci USA, 1996, 93: 7486 ~ 7490

14 Crameri A, Cwirla S, Stemmer W P. Construction and evolution of antibody-phage libraries by DNA shuffling. Nat Med, 1996, 2: 100 ~ 102

15 Perler F B. InBase, the intein database. Nucleic Acids Research, 2000, 28: 344 ~ 345

16 Otomo T, Ito N, Kyogoku Y, *et al*. NMR observation of selected segments in a larger protein: central-segment isotope labeling through intein-mediated ligation. Biochemistry, 1999, 38: 16040 ~ 16044

17 Kumagai I, Takata S, Miura K I. Functional conversion of the homologous proteins α-lactalbumin and lysozyme by exon exchange. Proc Natl Acad Sci USA, 1992, 89: 5887 ~ 5891

18 Tordai H, Banyai L, Patthy L. The PAN module: the N-terminal domains of plasminogen and hepatocyte growth factor are homologous with the apple domains of the prekallikrein family and with a novel domain found in numerous nematode proteins. FEBS Lett, 1999, 461: 63 ~ 67

19 Chang J Y, Monroe D M, Stafford D W, *et al*. Replacing the first epidermal growth factor-like domain of factor IX with that of factor VII enhances activity *in vitro* and in canine hemophilia B. J Clin Invest, 1997, 100: 886 ~ 892

20 Hahn D, Illisson R, Metspalu A, *et al*. Human N-benzoyl-L-tyrosyl-p-aminobenzoic acid hydrolase (human meprin): genomic structure of the alpha and beta subunits. Biochem J, 2000, 346: 83 ~ 91

21 Takahara K, Lee S, Wood S, *et al*. Structural organization and genetic localization of the human bone morphogenetic protein 1/mammalian tolloid gene. Genomics, 1995, 29: 9 ~ 15

22 Rossignol M, Gagnon M L, Klagsbrun M. Genomic organization of human neuropilin-1 and neuropilin-2 genes: identification and distribution of splice variants and soluble isoforms. Genomics, 2000, 70: 211 ~ 222

第 7 章　抗体酶定向分子进化

7.1 引　　言

抗体酶(antibody enzyme, abzyme, Abz)又称催化抗体(catalytic antibody),其本质是一类具有催化活性的免疫球蛋白,在其可变区赋予了酶的属性[1~4]。

酶与抗体具有相似性,都是具有高度特异性和亲和性的分子。几乎针对任何抗原,免疫系统都能生成相对应的抗体。科学家试图利用酶与抗体的相似性及免疫系统的特性制备具有催化活性的抗体。1969 年,Jencks 根据 Pauling 的化学反应过渡态理论提出用化学反应的过渡态类似物作半抗原(hapten)去免疫动物可能会产生具有催化活性的抗体,即抗体酶。1975 年单克隆抗体制备技术的出现为抗体酶的开发铺平了道路。1986 年,Lerner 和 Schullz 同时报道了具有催化化学反应活性的抗体酶。此后,以各种过渡态类似物作为半抗原成功地开发出天然酶所催化六类酶促反应的抗体酶。目前抗体酶催化的化学反应已有 80 余种。除酯、羧酸和酰胺键的水解反应外,还有酰胺形成、光诱导裂解和聚合、酯交换、内酯化、克来森重排反应、金属螯合、环氧化反应、氧化还原反应、化学上不利的环化反应、肽键形成反应、脱羧反应及过氧化反应等。这些抗体酶催化的反应专一性相当于或超过酶反应的专一性,催化速度有的可达到酶催化的水平。但一般来说,抗体酶催化反应的速度比非催化反应快 $10^2 \sim 10^6$ 倍,仍比酶催化反应速度慢 $10^2 \sim 10^4$ 倍。因此,开发制备高活力抗体酶的方法仍是科学家的奋斗目标。

抗体多样性是由基因重排,突变和选择/筛选产生的。体内迅速进化的这种原理和免疫系统的结合,使它们成为进化蛋白质设计方法的天然靶。随着分子生物学方法的发展,在 *E. coli* 和酵母中重组 Fab 或 scFv 的生产,抗体的临床和分析应用成为可能。另一个关键的发现是噬菌体展示技术的开发,它允许抗体片段展示在噬菌体的表面,由抗体库中筛选出所要求的抗体。紧接着发展了细菌和酵母表面展示技术和体外核糖体展示技术,并用于抗体的展示。以合成的多肽链—编码 RNA 共价融合为基础的技术展示抗体有很好的前景。这些展示技术同结合检验(例如生物素-链霉结合素系统或抗体的选择性标记)组合,可以迅速地筛选非常大的天然的、合成的或工程抗体库,可以迅速筛选 10^9 以上的结合蛋白质突变体全库。结合蛋白质的所有重要性质,例如特异性、亲和性、稳定性和折叠都可以通过定向进化改进。

7.2　抗体酶定向分子进化策略

7.2.1　突变策略

7.2.1.1　体外突变[2]

用易错 PCR(error-prone PCR)方法,可在噬菌体抗体的可变区基因或 CDR 区域的碱基中,产生随机点突变(random point mutation),几次选淘之后所筛选出来的抗体其亲和力提高 4 倍以上。但是,应用易错 PCR 方法,需要预先对所筛选出来的抗体基因进行序列分析之后再设计引物。有一种称作 CDR 巡查(CDR walking)的方法用于在体外引入抗体基因突变。该方法无需知道所筛选的抗体基因序列,而只根据 Kabat 等所提供的重链 CDR 信息,采用重叠 PCR 诱变技术在重链 CDR1 的 32 位点和 CDR3 的 96 ~ 99 位点引入碱基多样性。经筛选后,其抗体亲和力提高 10 倍以上。

7.2.1.2　体内突变[2]

大肠杆菌菌株 mutD5,其 DNA 聚合酶 Ⅲ 全酶中的 ε 亚基(epsilon subunit,该亚基具有 3′ ~ 5′外切酶校正活性)的基因 *mutD* 发生了突变,丧失了其校正功能,从而在该菌株复制时增加其体内基因突变频率。该菌株的 *LacI* 基因内易发生单个碱基取代,其中 A·T→G·C→A·T 转换频率相同,A·T→T·A 的碱基颠换要比 A·T → C·G 的碱基转换普遍且频率高。该菌株的自然突变频率比野生型高 10^3 ~ 10^5 倍,其中在 Lac 操纵子区的 + 6 位有一个碱基 A·T→G·C 转换的热点,其突变频率可高达 1.2×10^{-3}。

在应用体内抗体突变时,可将携带有 Fab(或 scFv)的噬菌粒转入 mutD5 等高突变菌株,也可将重链的可变区基因或轻链的可变区基因通过该突变株,甚至用 CDR 区片段克隆到该菌株的高突变热点。经过一代(或几代)繁殖后,从该菌株提取并切下来抗体基因片段,再连接到噬菌粒中,利用噬菌体筛选高亲和力抗体。

7.2.1.3　链改组[2]

由于在库容量不太大的情况下,B 细胞原有重轻链组合容易丢失,从原始库中筛选出来的抗体亲和力偏低。所以,人们试图从原始库中已筛选出来的可变区基因再重新组合,使其进一步多样化,从而提高其亲和力。这种再重新配组的方法称

为链改组(chain shuffling)。同时链改组法也有助于研究抗体结构与功能以及亲和力变化之间的关系。

链改组有几种情况,重链不变、改组轻链、使该重链与原始库中所有的轻链组合,或与已筛选出的几(或几十)株抗体中的轻链组合;轻链不变、改组重链、其组合方式与以上轻链组合方式相同。实践证明,链改组法是一条切实可行的提高抗体亲和力的途径,改组轻链已得到了增加亲和力 20 倍的抗体,改组重链提高亲和力 15 倍。

7.2.2 组合抗体库策略[1]

各种组合策略已用于产生高度多样性的抗体库,可以由库选择出高亲和性和新特异性的抗体。一种策略是高度可变 CDR 顺序(人 B 细胞的 cDNA)用 PCR 扩增,并随机重组到人免疫球蛋白基因单一主骨架 6 个可变区,在细菌中很好地表达并展示在噬菌体上。用该库已选择出各种不同蛋白质,肽和小分子抗原的特异 scFv 抗体。

组合抗体库(combinatorial immunoglobulin library)技术要点:用逆转录-PCR 技术从淋巴细胞克隆出抗体轻链基因组分(repertoire)和重链 Fd 段基因组分,将二者分别组建到表达载体 Lc2 和 Hc2 中,得到的轻链基因和 Fd 段基因随机重组于一个表达载体中,形成组合抗体库。所得到的抗体库经体外包装后感染大肠杆菌,铺板培养,每个感染了噬菌体颗粒的大肠杆菌细胞由于噬菌体的增殖而裂解,所释放的噬菌体再感染周围的大肠杆菌细胞,在培养皿细菌生长层内产生噬菌斑,同时表达的 Fab 片断也释放于噬菌斑内,将噬菌斑转印到硝酸纤维素膜上,可以用标记有过氧化物酶的抗原筛选得到产生特异性抗体的克隆,获其 Fab 段的基因。这个方法较细胞融合杂交瘤技术制备单抗有明显的优越性:①省去了细胞融合步骤,省时省力,可避免因杂交瘤不稳定而要反复亚克隆的烦琐程序;②扩大了筛选容量。用杂交瘤技术一般筛选能力在上千个克隆以内,而抗体库可筛选 10^6 以上个克隆;③用此技术可直接克隆到抗体的基因,既克服了杂交瘤分泌抗体不稳定而丢失的弱点,又便于进一步构建各种基因工程抗体;④用此法得到的抗体可以在原核系统表达,降低了制备成本;⑤构建抗体库时,轻链和重链可变区基因在体外随机组合,可产生体内不存在的轻重链匹配,有可能得到新的特异性单链抗体。单链 Fv 的优点是,分子质量小(2.6kDa),便于结构分析,同时提高了穿透组织的能力;此外单链 Fv 大大降低了抗体的免疫原性,减小了治疗中的不良反应;更重要的是,单链 Fv 可在 *E. coli* 中表达,为催化抗体的大规模应用奠定了基础。

7.2.3　噬菌体抗体库[1]

噬菌体抗体库技术为抗体酶的制备提供了更好的方法。Chen 等[7]用烷基磷酰胺(phosphoramidate)作半抗原免疫小鼠后,从中筛选出 22 个能同抗原结合的克隆,纯化后发现其中的 3 个克隆有催化活性,表征了其中的一个克隆,发现其动力学行为符合米氏动力学(K_m = 115μmol, K_{cat} = 0.25/min)。这是第一个从抗体库中筛选出的催化抗体。

虽然用噬菌体抗体库技术可有效筛选具有亲和力的抗体,但仅靠亲和特性筛选抗体酶还有困难。这是因为具有亲和力的抗体并不都有催化活性,实际上,具有催化活性的抗体只占结合抗体中的少数。另外,用 *E. coli* 表达的 Fab、单链 Fv,其表达量一般在 μg/L 到 mg/L,要得到足够量的样品,用于动力学分析也有困难。

为了减少筛选工作量,将酶的催化机制引入抗体库直接选淘,筛选步骤如图 7.1 所示。第一步用一般洗涤洗去非特异结合的噬菌体;第二步用 pH2.2 的盐酸缓冲溶液洗去非共价结合的噬菌体;第三步用 20mmol DTT 洗出共价结合的噬菌体,用于进一步扩增。经过 5 轮筛选后,随机挑取 10 个克隆测试,发现它能催化硫酯的水解反应。对其中的一个进行了研究,发现它能催化硫酯的水解反应,反应遵循简单的饱和动力学,K_{cat} = 0.030/min, K_m = 100μmol。

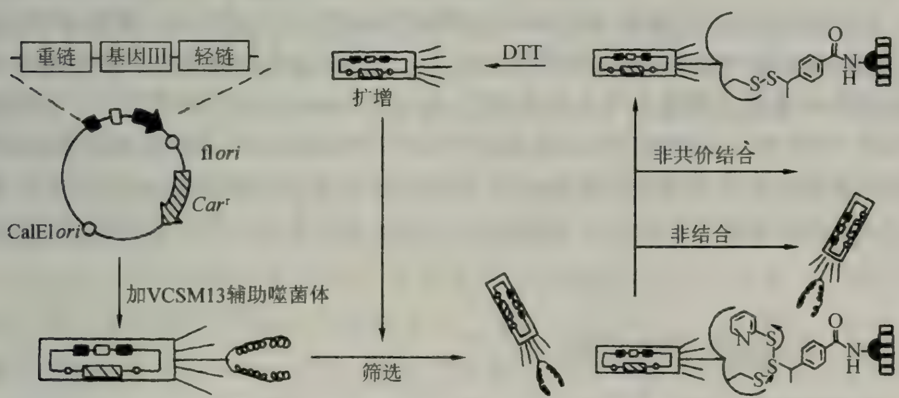


图 7.1　用 Fab 片断的噬菌体展示库筛选含未配体半胱氨酸的抗体结合部位的示意图

DDT,二硫苏糖醇; *Car*^r,羧苄青霉素抗性基因

7.2.4　合成抗体库策略

Knappik 等[5]合成了人 scFv 库(2 × 10^9 成员),包括 95% 以上人抗体多样性,在

所有各种组合中通过组合 7 个重链和轻链骨架的一致顺序，然后引入完全随机的 CDR3 盒。为了增加该库的多样性，在 6 轮易错 PCR 之后进行体外核糖体展示和选择，对牛胰岛素高亲和抗体较之合成库选出的抗体富集 40 倍（K_d = 82pmol/L）。选出的 scFv 有 2～9 点突变，并由不同的骨架组成。

7.2.5 结合荧光素 scFv 抗体库[6]

结合荧光素 scFv 抗体库的核糖体展示是通过一轮易错 PCR 继之几轮 DNA 改组产生的，并且可以选择出改进亲和力 30 倍的抗体（K_d = 0.04nmol/L）。在第二个实验中，在体外翻译和蛋白质折叠步骤逐渐增加还原的条件下经五轮选择产生突变体 scFv（scFv 含有两个保守的二硫键），不仅在体外还原条件下更有效地折叠，而且表明在 *E. coli* 外周质功能表达提高 4 倍，并且可以在还原胞质中功能表达。在两个实验中，选择突变体的累积突变分布在整个顺序，形成选择功能的各种顺序。为了改进折叠抗荧光素结合抗体的稳定性，最初是把它的结合突环嫁接到很好折叠的人抗体骨架上[7]。为了进一步改进它的稳定性，在增加温度或盐酸胍胁迫的情况下，应用几轮 DNA 改组和噬菌体展示库，最好的突变体（仅两点突变）所表现出的热力学稳定性达 4cal/mol（1cal = 4.1868J），其中一个定位在一个 CDR，并且结合常数增加 20 倍。

抗体的治疗应用，例如，癌症的治疗需要抗体优先不可逆结合。哺乳动物免疫系统通过体细胞突变产生高亲和性抗体，这个过程称为亲和性成熟。然而，体外亲和性成熟的遗传上限阻止具有解离常数 K_d < 0.1nmol/L 抗体的产生[8]。通过体外进化，完整的 scFv 顺序仅仅 4 轮易错 PCR 和 DNA 改组，继之酵母表面展示和展示的突变体库最佳化动力学筛选，产生最佳突变体（解离速率降低 1000 多倍）。在各种选择的高亲和性突变体中，所鉴定的一些一致突变发生在体细胞超突变过程不可及区。

7.2.6 T-细胞受体策略

T-细胞受体（scTCR）是产生多样性另一组免疫结合蛋白质（不涉及体细胞突变），成为经得起检验的生物学设计途径[9]。与 scFv 抗体比较，当在 *E. coli* 表达时，重组 scTCR 不太稳定且可溶性差。scTCR 随机突变，继之酵母表面展示，选择便于展示且保持抗原亲和性的突变体。有趣的是，许多鉴定的突变改变了 α-链和 β-链界面上特异可变位点的残基，这些残基是天然存在于许多更稳定抗体的可变区。然后，展示的最好的 3 个突变体同野生型 scTCR 一起混合并进行易错 PCR，这样产生的库展示在酵母表面，为提高耐热性和改进在酵母中的分泌进行平行

FACS 筛选。在 3 轮富集后获得稳定性明显改进的克隆，可以在 65℃时（与出发突变体在 45℃比较）分泌量增加 20 倍（2000mg/L）。与抗体不同，TCR 在体内亲和性成熟过程中不经历体细胞突变，报道的 TCR 亲和性（$10^5 \sim 10^7$L/mol）低于抗体（$10^5 \sim 10^{10}$L/mol）。然而，scTCR 的体外进化可使亲和性增加约 100 倍（$K_d \approx 9$nmol/L）[10]。

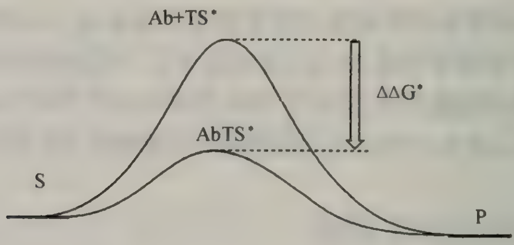
7.2.7　提高催化活性策略

Takahashi 等[11]开发出提高抗体酶催化活性（K_{cat}）策略——定向进化抗原-抗体识别以改进抗体酶催化活性的体外策略。改进的突变体催化活性提高 6～20 倍（见 7.4 节）。

7.3　抗体酶定向分子进化筛选/选择策略

7.3.1　结合、催化和免疫组分策略[12,13]

各种免疫球蛋白是免疫系统的蛋白质。它们的多样性主要在组成抗原结合部位的互补决定区（CDR），是 B 细胞分化时基因顺序重组产生的。虽然多样性在抗体成熟时可以通过体细胞随机突变扩增，但理论上总的有效多样性不大于 10^{12}。因此，从半抗原（反应过渡态类似物或反应性化合物）开始搜索抗体酶。实际上是由可能的抗体库中选择抗体酶，这样，选择抗体酶首先考虑结合能。抗体酶是典型地对着小分子半抗原（化学反应的过渡态类似物）提升。在最好的情况下，人们可以预期抗体-抗原相互作用全部有效的结合能将转为过渡态稳定化，以 $\Delta\Delta G^*$ 表示，它直接与过渡态解离常数 K_{TS}相关（图 7.2）。这样，典型的抗体-抗原复合物解离常数在 nmol/L（$K_d = 10^{-9}$mol/L）范围，可以转化为反应过渡态的解离常数（$K_{TS} = 10^{-9}$mol/L）。对于一种底物的反应，具有典型的 Michaelis-Menten 常数 $K_m = 10^{-3}$ mol/L，那么特异的速率加速可以预期是 $K_{cat}/K_{uncat} = K_m/K_{TS} = 10^6$，相应于降低过渡态能量大约 8.4kcal/mol。其次，考虑“抗体催化”反应涉及所用半抗原的实际结构。免疫库的抗体实践上是可及的，或者用单克隆抗体技术，或者应用噬菌体展示技术。在上述情况下，免疫系统在含水环境通过结合亲和力进行操作。它是通过疏水力驱动，用脂肪族或芳香族取代。设计半抗原最简单的和最有效的途径是带有芳香基并相应于反应过渡态类似物中的底物。底物中的芳香基也有利于反应产物的动力学分析，最常采用的是 UV 或荧光分析。



$$AbTS^* \underset{}{\overset{K_{TS}}{\rightleftharpoons}} Ab+TS^* \quad K_{TS}=\frac{[Ab][TS^*]}{[AbTS^*]}$$

图 7.2 经过渡态 TS* 或抗体过渡态 AbTS*，抗体催化和未催化底物 S 转化为产物 P
抗体的催化作用是通过 ΔΔG* 给出，其相应于过渡态解离常数

7.3.2 理论检验策略

7.3.2.1 筛选的敏感性

抗体酶经常以活性系统开始操作，其背景表明反应的非催化水平。抗体酶的底物饱和动力学遵循米氏方程。为了使底物浓度接近零，催化/非催化反应比率增加到最大(图 7.3)。同时反应的绝对速率趋向零。这样，抗体催化作用在最低可能的底物浓度(产物形成仍然是可检测)将是最好检验的。应用高底物浓度对于饱和底物结合部位可能是必需的。结合部位的饱和只使反应有意义，但不表明任何未催化成分。

催化剂检验敏感性第二个方面是涉及检试溶液中潜在的有效催化剂浓度。在最适条件下，底物浓度在 K_m 以下观测的反应速率由下式给出

$$v_{obs} = v_{uncat} + v_{cat} = K_{uncat} \times [S] + (K_{cat}/K_m) \times [Ab] \times [S]([S] \ll K_m)$$

$$v_{cat/uncat} = [(K_{cat}/K_{uncat})/K_m] \times [Ab] = [Ab]/K_{TS}$$

考察已发表的抗体酶的资料表明，许多抗体酶具有过渡态解离常数 K_{TS}大约是 10^{-7}mol/L，其相应于特异的速率提高，$K_{cat}/K_{uncat} = 10^3$，对于每个结合部位分子质量约 70kDa 抗体，最低筛选浓度 2μg/ml = 2mg/L 是检验活性水平所需要的。在杂交瘤细胞培养中单克隆抗体的浓度一般适于分析。相反，在非最佳条件下重组子表达系统一般达不到这样的抗体浓度。

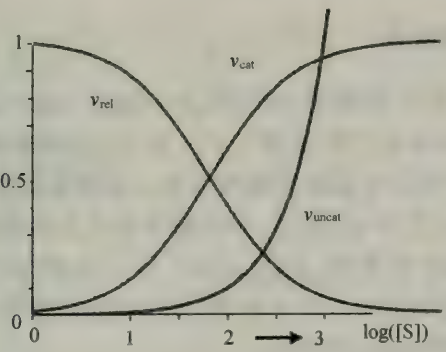


图 7.3　催化速率 v_{cat}，未催化背景反应 v_{uncat}和两者 v_{rel}的比率作为底物浓度的函数的对数表示图解
一种底物反应遵守米氏预平衡结合模型

7.3.2.2　筛选的选择性

筛选的选择性涉及对观察所要求的催化反应速率的影响。给定的抗体样品反应速率超过缺乏抗体的样品的未催化反应，这并不意味观察效应实际上是来自抗体的结合部位。在抗抗原抗体的情况下，假设催化部位紧密结合抑制剂在免疫抗原情况中是有效的，观察催化效应受半抗原抑制可以在筛选时立即检试。这是确定催化作用来源的一种简单的、必需的条件。催化作用也应当根据多重转换率来证明。关键问题之一是遇到抗体被产物抑制。有效催化作用（低的 K_{TS}）一般是通过观察对半抗原高亲和力（低 K_i）。在许多抗体酶已观察到催化效率与半抗原结合之间的相关性。

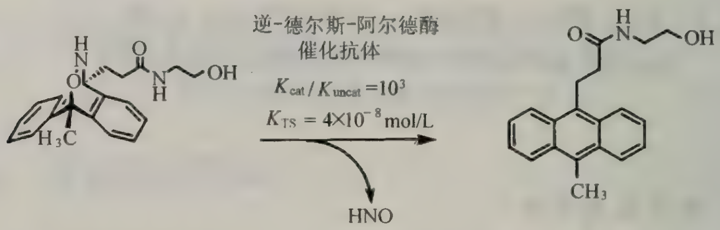
紧密结合半抗原的观察进一步证明抗体催化作用。因为催化活性可以清楚预计抗体的结合部位数。定量抑制作用可有效地排除可能的酶污染。当研究典型的“自然发生反应”（abiological reaction）时，酶污染的危险通常是低的。从酶设计的观点来看这是最有趣的。当研究典型类酶反应时，酶污染可能是很成问题的。抗体催化作用最终证据是必须观察到活性。

7.3.2.3　筛选实践[14,15]

检验的敏感性和选择性一般可以认为是发现抗体酶的必要参数。应用荧光和生色反应筛选杂交瘤产生的抗体，原则上可以迅速和简单地检验任何数量样品的催化作用。

1. 逆-狄尔斯-阿德尔抗体酶

双环前药物 **1** 逆-狄尔斯-阿德尔反应(retro-Diels-Alder reaction)释放硝酰(基)(一种调节 NO 的前体)和蒽产物 **2**(图 7.4)。这个反应很容易根据荧光测定,因为底物是非荧光的,而产物 **2** 在蓝区有强烈荧光。这种非生物逆-狄尔斯-阿德尔抗体酶在 pH4 ~ 10 之间似乎与 pH 无关,并对缓冲液或共溶剂效应完全不敏感。同时,这个反应有可测的自发速率($K_{uncat}=10^{-5}/s$)。这样,该反应是在抗体催化的有效参数之内。最重要的是该反应不被蛋白质(例如清蛋白、抗体或酶)非特异催化,因此可以直接在细胞培养中测定。



1　　　　　**2**

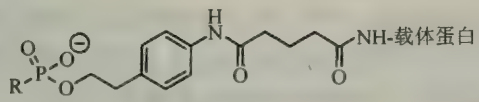
图 7.4　抗体催化逆-狄尔斯-阿德尔反应

通过荧光高通量筛选,能够筛选 14 000 以上杂交瘤细胞培养样品,其是由鼠对反应的过渡态类似物免疫产生的,并分离出 8 个抗体酶,$K_{cat}/K_{uncat}=10^2\sim10^3$。这些速率提高是在典型抗体酶范围内,$K_{TS}$范围在 0.1 ~ 0.9μmol/L 之间。这种高通量筛选甚至弱活性抗体($K_{cat}/K_{uncat}=100$)也容易在细胞培养中检出。活性最高的抗体酶已经进行了详细地研究,其作用机制根据 X 射线晶体结构已确立。

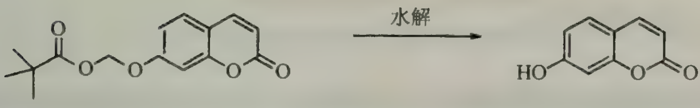
2. 新戊酸酶活性抗体酶

酯酶可以存在于细胞培养样品中,并可以观察其催化作用的发生。对磷酸半抗原 **3—5** 免疫,用荧光底物 **6** 作为检试底物,在水解时释放强荧光 7-羟基香豆素(图 7.5)。虽然新戊酰氧甲基(POM)衍生物 **6** 在细胞培养中表现出很小的非特异水解作用,但是它是潜在的酯酶底物。应用细胞培养进行第一轮免疫和筛选,用 ELISA 检验强的半抗原结合,并通过加入半抗原抑制催化作用。该实验产生五个单克隆抗体细胞系,表现出特异的半抗原抑制水解。然而抗体的纯化表明没有与蛋白质 G 抗体亲和柱结合,最大可能是由酯酶污染造成的。通过半抗原抑制催化作用更弱(10^{-5}mol/L),并且用 ELISA 检验半抗原对抗体没有强的亲和力,这进一步证明催化活性不是来自抗体。

根据这些实验,筛选杂交瘤细胞培养样品可以应用抗体预纯化操作(图 7.6)。



3 (R=t-Bu)
4 ($R=CH_2Cl$)
5 ($R=C_6H_5$)



6　　　　**7**

图 7.5　抗体催化新戊酸氧甲基(POM)酯的水解

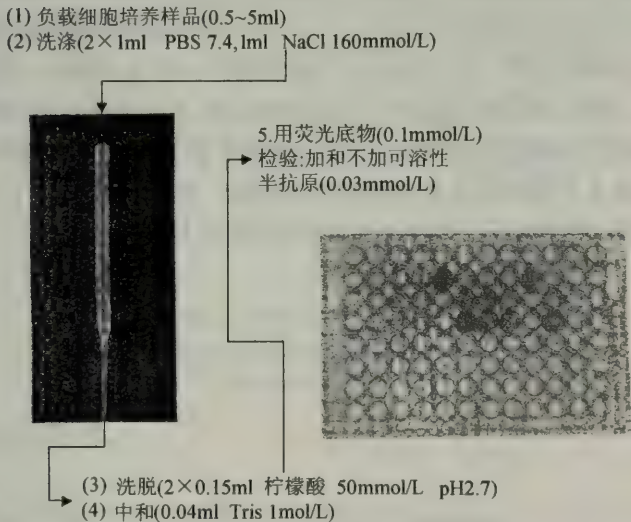


图 7.6　检验杂交瘤细胞培养样品的抗体预纯化操作

0.5 ~ 5ml 细胞培养样品中的免疫球蛋白通过很小量的蛋白 G 凝胶(简单地用 Pasteur Pipetle 填充)选择性地结合。凝胶用中性缓冲液洗涤,然后抗体用小体积的酸性缓冲液选择性地洗脱。洗脱样品被中和,通过比较半抗原存在或缺乏情况下荧光底物 **6** 反应速率可以特异的检验半抗原抑制催化。

蛋白质 G 预纯化操作可以在 2h 内手工操作纯化 40 个样品。催化作用筛选是在大约 5ml 细胞培养体积的杂交瘤样品中进行的。选择 60 个细胞系,检验了半抗原抑制催化并进一步培养。所有的抗体成功地两次亚克隆。7 个已确定在表达和纯化之后半抗原抑制催化。不仅在第一次克隆鉴定催化作用,而且在第一次和第

二次亚克隆之后,催化检试也是必需的。的确,来自初始催化杂交瘤的亚克隆常常在 ELISA 发现阳性,但催化作用却是阴性,而其他 ELISA 是阴性,催化作用也是阴性。分离出的新戊酸酶抗体酶是第一个能够水解新戊酸酯的抗体。

7.4 抗体酶定向分子进化筛选/选择方法

7.4.1 直接筛选[1]

在微生物的自然代谢途径中,尿苷酸的合成是由乳清酸磷酸核糖转移酶(OPRTase)及乳清酸脱羧酶(ODCase)依次完成。如果一个微生物是 ODCase 缺陷型,那它必须含尿嘧啶磷酸核糖转移酶(Ura-PRTase)将尿嘧啶直接转变成尿苷酸(UMP)。如果设计的抗体酶具有催化乳清酸脱羧作用,则能弥补微生物的缺陷,在不含尿嘧啶的培养基中也能生长(图 7.7)。为了获得乳清酸脱羧抗体,Smiley 等合成了过渡态类似物半抗原[图 7.8(b)]。将此抗原与 KLH 偶联后,免疫小鼠,构建噬菌体抗体库,感染 ODCase 缺陷型大肠杆菌后,在不含尿嘧啶的培养基上生长。这样只有含催化乳清酸脱羧的抗体基因的菌株才能长出克隆。结果从 16 000 个转化体中筛选出 6 个具有催化乳清酸脱羧的克隆。他们把其中的 1 个单克隆抗体转变成单链抗体,并在大肠杆菌中表达,结果证明也有明显活性。

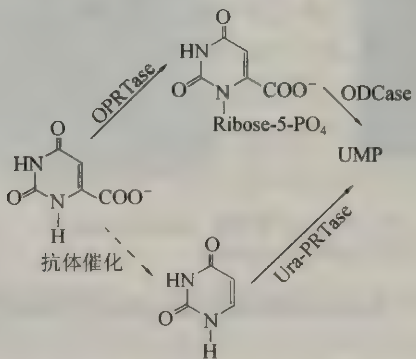


图 7.7 尿苷酸(UMP)生物合成途径

上部途径包含天然酶 OPRTase 和乳清酸脱羧酸酶(ODCase);缺少 OPRTase 和/或 ODCase 的突变株,在抗体酶(具有乳清酸脱羧酶活力)和天然酶 Ura-PRTase 存在下,仍可按下部途径合成 UMP

7.4.2 化学筛选[1]

应用筛选试剂可从抗体组合库中通过化学筛选获得糖苷酶抗体。半抗原 A

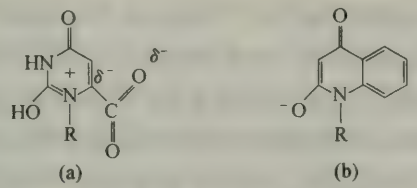


图 7.8　诱导乳清酸脱羧抗体的半抗原设计

(a) 乳清酸脱羧的过渡态；(b) 设计的半抗原

中糖苷键水解产生醌甲基化物 B，它能共价捕获抗体库中的具有糖苷酶活力的 Fab（图 7.9）。筛选出的 Fab 能催化水解 *p*-硝基苯基 β-吡喃半乳糖糖苷，速度加强比（ER）为 70 000，而且经典的过渡态类似物法所制备的最好的糖苷酶抗体，其 ER 值仅有 100。对于任何难解离的反应，只要它能通过反应中间体捕获抗体都可用该法筛选出高效催化剂（见 3.3 节）。

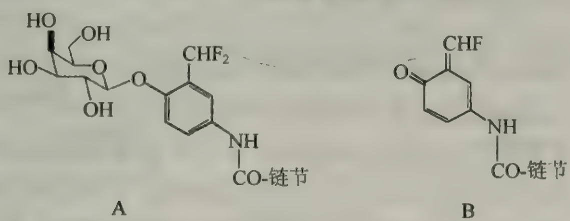


图 7.9　半抗原(A)用于化学筛选糖苷酶抗体片段 Fab；醌甲基化物(B)可捕获并显示具有催化性质的 Fab

7.4.3　遗传选择

遗传选择进化酶（改进催化活性、稳定性和改变专一性）是相当有效的。因此，为提高抗体的催化活性，开发遗传选择是相当有意义的研究工作。对遗传选择而言，必须确定活性酶/抗体表达生长优势的条件。对高催化活性的抗体酶而言，这种条件的确定相对容易，对于催化活性低的抗体酶而言，确定这种条件是困难的。然而，影响细菌生长速度的其他因子，例如毒性差别和抗体的表达水平可以完成遗传选择。Schultz 等[16]报道了抗体酶催化活性遗传选择的一种策略——*E. coli* 生长优势策略，生长优势是催化活性的直接结果。

开发遗传选择首先是选择营养缺陷型宿主和互补策略。一种很好的选择宿主应当容易操作，能够产生活性抗体并适合各种底物范围的选择。根据这些要求，Schultz 等采用 *E. coli* HDU78，由于其缺失分支酸合成酶基因（aroc）不能合成分支酸。*E. coli* HDU78［hdsFmcr13 araD139△（ara ABC-leu）7679△Lac × 74galUgalK rpsl

thi-aroc purF: Tn10]是由菌株 MC1061[hdsFmcr13 araD139△(ara ABC-leu)7679 △Lac×74 galUgalK rpsl thi-]通过一系列 P1 转导产生的。分支酸是莽草酸途径一个关键中间体,并且主要的代谢产物酪氨酸,苯丙氨酸,色氨酸,对氨基苯甲酸(PABA),对羟基苯甲酸和异分支酸的生物合成是需要的。原则上,这些营养成分的任何一种可以作为互补的基础。HDU78 能够表达活性形式的 Fab,产量 100~200μg/L。

开发遗传选择其次是挑选选择标记和反应系统。以 PABA 为基础互补应当是相当普遍的,因为化学反应范围(例如水解、消除、氧化还原等)可以与营养成分的释放偶联。此外,以 PABA 为基础的互补应当将催化活性转为可测的生长优势,因为维持 *E. coli* 生长所需要的 PABA 浓度非常低(1~10nmol/L)。为了检验抗体催化互补反应中 PABA 的效用,研究了醛缩酶抗体 38C2。该抗体酶底物谱广,特别是催化前体药物的活化,例如,**1** 经逆-3-羟基丁醛逆-迈克尔加成反应顺序(图 7.10)。这个系统对于 PABA 释放容易修饰。

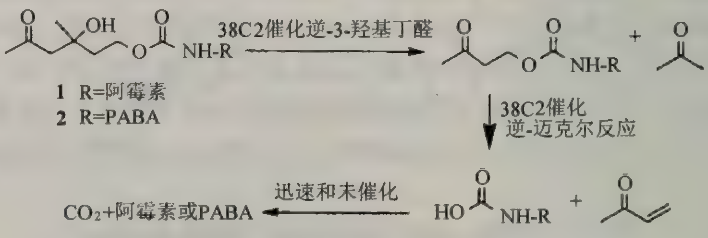


图 7.10 前药物 **1** 和底物 **2** 的逆-3-羟基丁醛-逆-迈克尔活化

涉及 38C2 催化 PABA 释放的一种遗传选择需要底物 **2** 不被任何的内源酶转化为 PABA。在培养介质中未催化反应的速率足够低,以提供可选择的窗口。底物 **2** 被合成(图 7.11),然后在各种浓度的 PABA 或 **2** 存在下比较 HDU78 的生长速率。为了测定活性抗体的表达提供生长优势并鉴定最佳生长条件,用表达 38C2 的细菌进行一系列模型选择,用表达 38C2 失活的突变体的细菌作为对照。

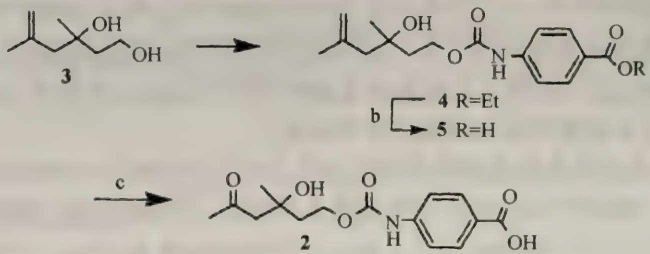


图 7.11 底物 **2** 的合成

总之,Schultz 等开发出以缺乏分支酸合成酶活性的 *E. coli* 营养缺陷型的互补为基础;抗体酶的新的遗传选择系统,通过一系列模型选择,成功地确立了抗体酶活性的表达提供生长优势的选择条件,即证实生长优势是催化活性的直接结果。该选择系统对催化活性转为可测的生长优势相当敏感。通过产生抗体的大组合库并进行重复几轮选择和 DNA 改组,可以鉴定出高活性的醛缩酶抗体。

7.4.4　动力学选择

应用噬菌体展示抗体库的体外进化是获得抗体酶的有效方法。Takahashi 等[26]提出了定向进化抗原-抗体识别以改进抗体酶催化活性的体外策略。

在酶的自然选择和抗体酶免疫选择之间是平行的。在免疫反应时,由多样性的抗体库中可以选择出 TSA 有亲和力的抗体,并且亲和力通过抗体基因的超突变(hypermutation)进一步最佳化。这样对底物过渡态的进化被认为是最大亲和力的进化。根据过渡态理论,抗体酶对过渡态亲和力强不一定反应速度最大(K_{cat})。因此,为了进化高催化活性的抗体酶,就得使亲和力的差异最佳化。然而,为了最佳化亲和力差异就得控制抗体对底物的过渡态和基态的识别。

Takahashi 等应用噬菌体展示技术,根据酶的进化动力学体外进化抗体酶。这样,初始的抗体酶噬菌体展示库对着新设计的 TSA 进行选择,使过渡态相对于基态的亲和力最佳化,这样提供具有改进反应速度(K_{cat})的突变体。进化的抗体酶同初始抗体的结构和动力学比较指出库的优点超过了通常的体内进化,体外进化对于产生新的催化剂以及提供检验酶进化动力学(dymamics)有很大的潜力。

作为检验体外定向进化的模型系统,Takahashi 等选择了抗体酶 6D9,它水解前体药物 **1** 产生氯霉素 **2**(图 7.12)。体外用 TSA3 免疫产生的抗体 6D9 作为抗体酶,其具有不同程度的水解活性。以前研究表明,在水解反应中 6D9 和其他的抗体酶,稳定由羟阴离子进攻酯羧基所产生的过渡态。6D9 的 X 射线结构分析和定位诱变确定了催化残基。应用动力学和结构信息,Takahashi 等合理地设计了 6D9 噬菌体展示库,对着新设计的 TSA4 选择,与体内方法所获得的初始抗体酶比较,可以产生高活性的进化抗体酶。

6D9 噬菌体展示库是通过 pComb3 系统修饰构建的,通过 PCR,制备了 6D9 轻链库,其中围绕催化 His 的 6 个氨基酸的密码子随机化(图 7.13)。为了相对于基态反应的过渡态的亲和力最佳化,设计了新的 TSA4。在 TSA4 中,化合物 **3** 的三氟乙酰基用于免疫,转为乙酰基。抗体酶的体外进化策略较之体内突出的优点是允许用一系列相应于 TSA 结构的抗原重复进行选择。

体外进化的单链抗体突变体,具有最高单价结合配体的亲和性。根据酵母表面展示抗体的最佳动力学筛选,在 4 轮亲和突变和筛选之后解离速率降低 1000 倍

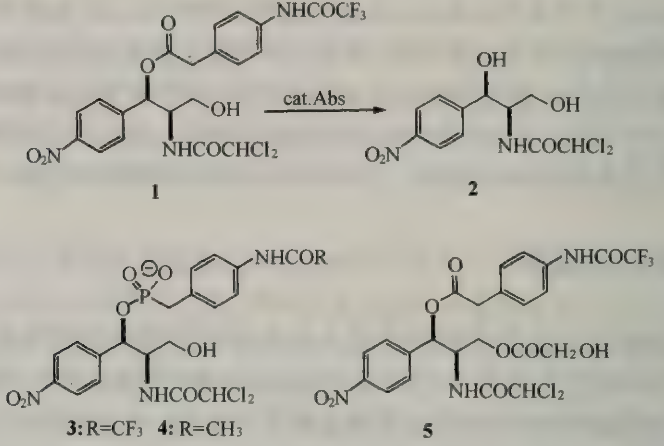


图 7.12 抗体催化前体药物活化

氯霉素单酯衍生物(**1**)被抗体酶水解产生氯霉素(**2**)。半抗原 **3**(水解的过渡态类似物)用于免疫产生抗体酶。半抗原 **4**(半抗原 **3** 的衍生物)用于噬菌体展示抗体酶的选淘,底物 **5** 用于动力学研究

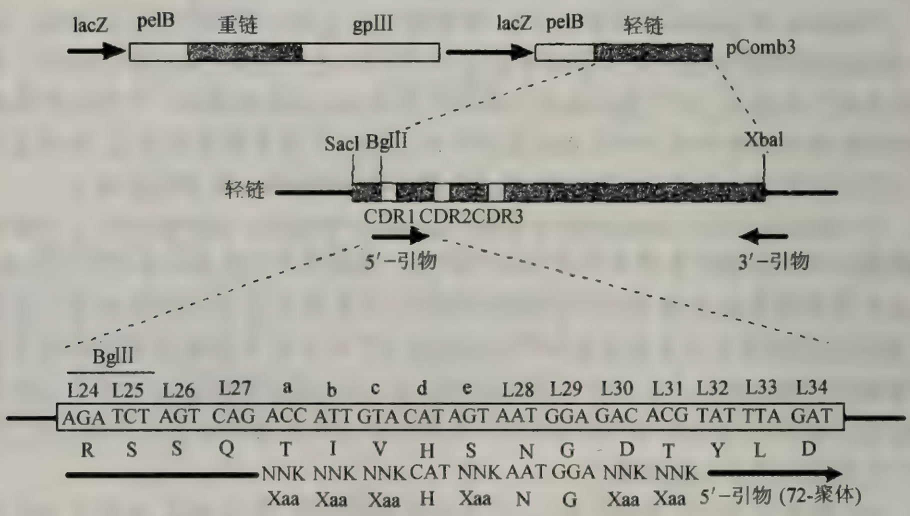


图 7.13 在 pComb3 载体中,噬菌体展示 6D9 轻链 CDR1 库的构建

以上。

目前,100 多种抗体临床试用于癌、病毒、自身免疫和其他疾病的治疗。高亲和性抗原的识别是所有治疗剂的核心问题,并且一般由哺乳动物免疫系统体内亲和性成熟来实现。由于亲和性是治疗应用的关键,通过定向进化已经产生较高亲

和性的抗体。抗体/半抗原识别已经通过结构、热力学、动力学、计算、光谱和突变方法进行表征。

7.4.5　应用反应化合物体外选择抗体酶[17]

酶催化作用的自然选择是改进功能的一种结果，从而提高进化过程中机体的适应度。相反，应用给定的化学转化反应配位的信息，根据与反应过渡态类似物或所设计的带电化合物的结合来选择抗体酶。如果在免疫系统内选择，可能将简单的结合转换为催化功能，从而有效的产生抗体酶，将这种策略称为反应免疫（reactive immunization）。反应免疫策略为高效抗体酶的制备提供了一种途径，用该途径制备的抗体酶的惟一特征是它们的底物范围广。

噬菌体展示抗体库已成为鉴定单克隆抗体的一种有效策略。在体外应用噬菌体抗体库以反应化合物进行化学选择可以获得抗体酶，并且不受动物来源或免疫反应的限制。

应用各种二酮衍生物的噬菌体库筛选策略，改变了醛缩酶抗体底物专一性和转换率。应用该策略制备的抗体酶，可有效地催化环己酮-3-羟基丁醛 **28**，因为 **28** 的逆-3-羟基丁醛反应与丙酮-3-羟基丁醛比较是相当低的（图 7.14）。噬菌体库通过对着 **30**-牛血清白蛋白（BSA）和 **1**-牛血清白蛋白选淘来选择。由这种选择获得的 Fab28 催化反式-**31**、顺式-**31** 和顺式-**32** 的逆-羟基丁醛反应。Fab28 的 K_{cat}值超过母体抗体 3～10 倍。此外，Fab28 同抗体 33F12 相似的 K_{cat}值催化丙酮-3-羟基丁醛。

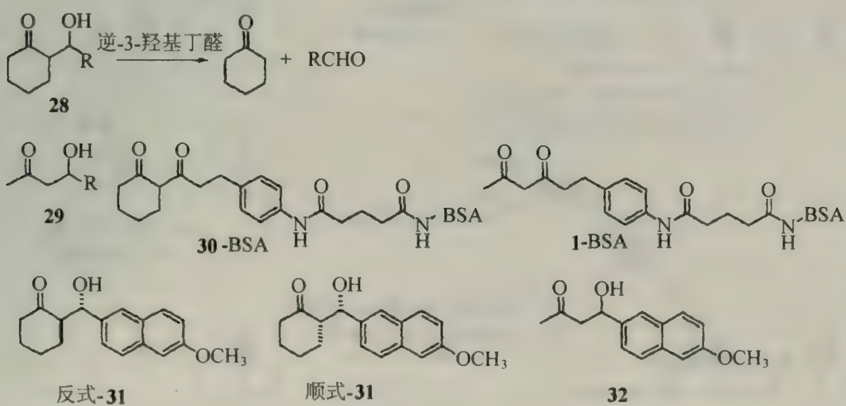


图 7.14　构建醛缩酶抗体的反应和化合物

7.4.6 催化抗-DNA 抗体的筛选方法

Mouratou 等[18]开发一种抗 DNA 水解抗体的微效价平板筛选方法——亲和连接寡核苷酸核酸酶检验(ALONA)方法(图 7.15)，所采用的 DNA 底物的 3′端生物

(a)

$d(T)_{40}$ 5′ (DIG) - TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT - Ⓑ 3′

subBV 5′ (DIG) -AAAAAAACCCCCCCATATAGCGCGTTTTTTT - Ⓑ 3′

(b)

$d(A)_{40}$:$d(T)_{40}$ 5′ (DIG) - TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT - Ⓑ 3′
3′ HO -AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA -OH 5′

ds subBV 5′ (DIG) -AAAAAAACCCCCCCATATAGCGCGTTTTTTT - Ⓑ 3′
3′ HO - TTTTTTTGGGGGGGTATATCGCGCAAAAAAA - OH 5′

图 7.15 (a)单股 DNA 底物的顺序。底物的 3′端用生物素Ⓑ标记，5′端用地高辛(DIG)标记；(b)相应标记双股 DNA 顺序

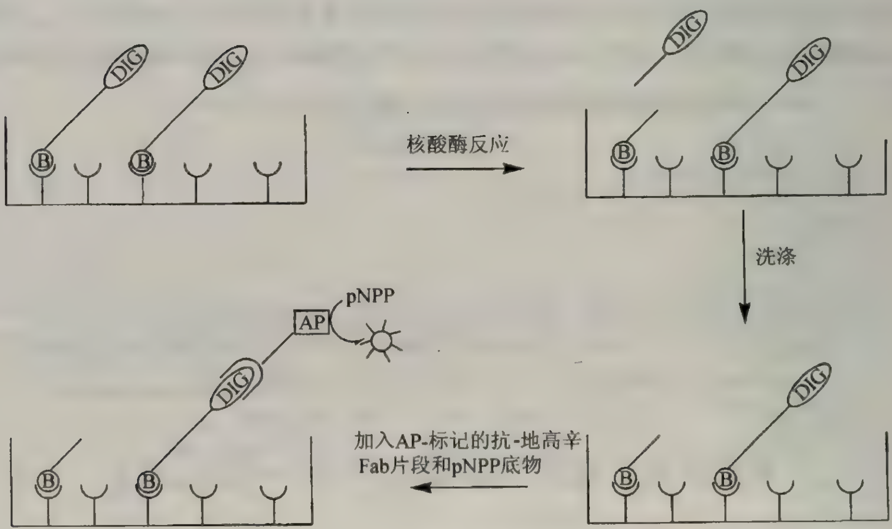


图 7.16 ALONA 微效价检验的图解

双标记的寡核苷酸底物固定在涂铺链霉结合素微效价平板上。反应允许在涂铺底物的孔中进行。然后用抗地高辛 ELISA 定量检验地高辛未分裂的 DNA。AP 表示碱性磷酸酶；pNPP 表示对硝基苯磷酸

素化(B),5′端地高辛配基标记(DIG)。这种底物专一地同涂铺链霉亲和素微效价平板孔结合并发生反应。未分裂的底物保留着高亲和标记,然后用酶标记的抗-地高辛抗体检验,通过测定 S1 核酸酶和 DNase I 活性及 EDTA 对反应的抑制作用评估检验效率(图 7.16)。ALONA 操作成功地用于大量的杂交瘤克隆的筛选。ALONA 方法对高通量筛选潜在的各种核酸酶和抗体酶是很有价值的。虽然该法是为筛选水解 DNA 抗体酶设计的,但它可以适用于检验其他类型底物分裂反应。

7.4.7　选择性感染噬菌体[4]

选择性感染噬菌体(selectively infective phage)利用不具感染力的 GIII 蛋白 N 端结构域缺失型噬菌体作为抗体库载体,将抗原与噬菌体缺失的 GIII 蛋白 N 端结构域融合,可以与抗原发生抗原抗体结合反应的噬菌体抗体因具有了缺失的 GIII 蛋白 N 端结构域,而重新获得感染力,通过这一方法可选择与半抗原结合的抗体。以此原理为基础,Gao 等进一步研究了将噬菌体感染力恢复与抗体催化活性直接相联系的抗体酶选择方法,如图 7.17。该方法以经抗体催化后可与抗体共价结合,并可与 GIII 蛋白结合的标记基团的双功能化合物作为选择底物,含催化抗体的失去感染力的噬菌体催化该化合物水解并与之结合,然后通过该化合物末端与

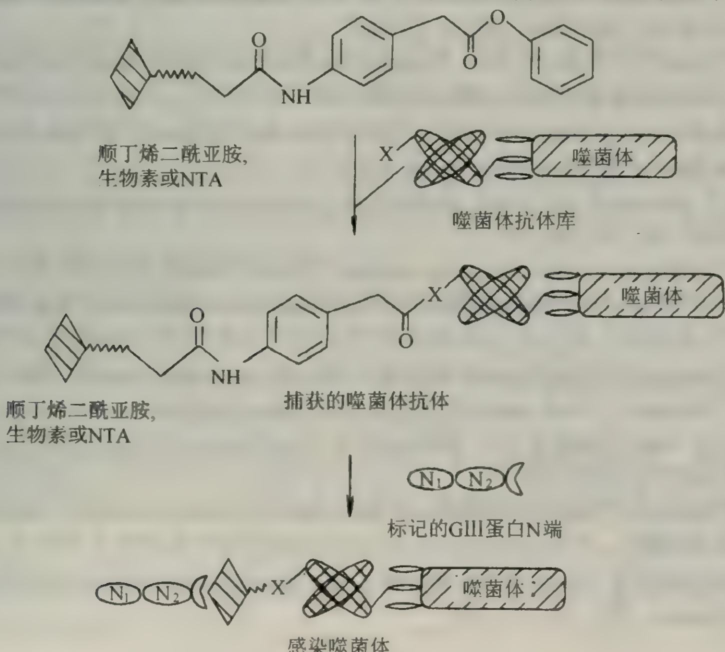


图 7.17　用标记的诱捕剂恢复噬菌体感染力选择亲和催化抗体

GⅢ 蛋白结合而使感染力恢复。

利用已鉴定的两个苯酯水解抗体对上述方法进行的研究发现，以共价机制水解苯酯的抗体 PCP21H3 非感染噬菌体的感染力可以得到有效恢复，而不以共价机制催化的抗体 PCP2H6 噬菌体的感染力不能恢复，说明了此方法具有较好的选择性。研究还发现，该方法具有很高的选出效率，富集因子可达 6×10^4，因而可能成为从较大的抗体库中选择催化抗体的有效方法。

参 考 文 献

1 罗贵民. 酶工程. 北京：化学工业出版社，2002

2 田波，李传昭，孙仑泉，王学. 分子进化工程. 北京：科学出版社，1999

3 沈倍奋. 分子文库. 北京：科学出版社，2001

4 甄永苏，邵荣光. 抗体工程药物. 北京：化学工业出版社，2002

5 Knappik A, Ge L M, Honegger A, *et al*. Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. J Mol Biol, 2000, 296: 57～86

6 Jermutus L, Honegger A, Schwesinger F, *et al*. Tailoring *in vitro* evolution for protein affinity or stability. Proc Natl Acad Sci USA, 2001, 98: 75～80

7 Jung S, Honegger A, Plückthun A. Selection for improved protein stability by phage display. J Mol Biol, 1999, 294: 163～180

8 Batista F D, Neuberger M S. Affinity dependence of the B cell response to antigen: a threshold, a ceiling, and the importance of off-rate. Immunology, 1998, 8: 751～759

9 Kieke M C, Shusta E V, Boder E T, *et al*. Selection of functional T cell receptor mutants from a yeast surface-display library. Proc Natl Acad Sci USA, 1999, 96: 5651～5656

10 Holler P D, Holman P O, Shusta E V, *et al*. *In vitro* evolution of a T cell receptor with high affinity for peptide/MHC. Proc Natl Acad Sci USA, 2000, 97: 5387～5392

11 Takahashi N, Kakinuma H, Liu L, *et al*. *In vitro* abzyme evolution to optimize antibody recognition for catalysis. Nat Biotechnol, 2001, 19: 563～567

12 Reymond J L. Detection strategies for catalytie antibodies. J Immunol Methods, 2002, 269: 125～131

13 Stevenson J D, Thomas N R. Catalytic antibodies and other biomimetic catalysts. Nat Prod Rep, 2000, 17: 535～577

14 Wahler D, Reymond J L. High-throughput screening for biocatalysts. Curr Opin Biotechnol, 2001, 12: 535～544

15 Wahler D, Reymond J L. Novel methods for biocatalysts screening. Curr Opin Chem Biol, 2001, 5: 152～158

16 Gildersleeve J, Janes J, Ulrich H, *et al*. Development of a genetic selection for catalytic antibodies. Boorg Medchem Lett, 2002, 12: 1691～1694

17 Tanaka F, Barbas C F. Reactive immunization: a unigue approach to catalytic antibodies. J Immunol Methods, 2002, 269: 67～79

18 Mouratou B, Rouyre S, Guesdon J L. A method for the detection and Screening of catalytic anti-DNA antibodies, J Immunol Methods, 2002, 269: 147～155

第 8 章　蛋白质组、代谢途径、病毒和细胞的定向进化

8.1 引　言

20 世纪后 50 年的生命科学基本上是围绕基因和蛋白质这条主线展开的。基因组计划的提出和实施导致了基因组学(genomics)的诞生,它是以生物体内全部基因为对象进行整体性研究。基因组学的研究已经取得了传统分子生物学难以取得的成就,生命科学进入了"后基因组时代"。在基因组学的推动下产生了蛋白质组学(proteomics),它研究基因组表达的全部蛋白质及其活动规律,即在细胞中蛋白质整体水平上进行研究。随着基因组学和蛋白质组学研究的深入,科学家认识到生命是一个复杂的非线性系统,它有自组织和自组装的特性。一些简单的分子组装成生物大分子以后就会出现某种新功能,这种功能难以从形成其物质基础推导出。

细胞是生命的基本单位,有着各种不同的生命活动,这些活动不是孤立的,而是相关联和高度协调的,形成细胞活动的网络结构。多种相互作用的分子构成复合结构,它是执行细胞功能的基本单位,称为"模块"(module)。一个功能模块是一个独特的整体,不同于其他模块。模块不是静态,而是动态的,不同时间有不同的分属。这就是说生命科学的研究的趋向似乎是个体性→整体性→复杂性,不同时期都将产生新概念,新方法和新学科。

生物进化的理论指导生命科学各个层次的研究以及各分支学科体系的建立。同时,进化论也随着生命科学各学科的发展而进化。无论是人或自然,还是一种理论或学科本身,都处于不断的进化之中。时至今日,进化论不再仅仅是一种理论,而是一门预见性科学,并逐渐成为一门应用科学。

定向进化的实质是达尔文进化论在分子水平上的延伸和应用。现在定向进化已由基因和蛋白质分子向基因组和蛋白质组、代谢途径和病毒甚至向细胞方面努力延伸。本书前 7 章都是围绕核酸和蛋白质定向进化展开的。本章介绍定向进化在基因组学、蛋白质组学、代谢途径、病毒和细胞方面的应用,但是应该指出这方面的工作还刚刚开始。

8.2 结构基因组定向进化的策略和方法

结构基因组学应该有传递三维结构信息的宏伟目标,然而至今只有小部分天然蛋白质适于结构测定。这是因为存在一些瓶颈问题,例如表达量少、聚集、错误折叠、稳定性差以及结晶困难等。一般通过它们的天然同源物紧密相关的可溶性高表达蛋白质来解决这些瓶颈问题。

结构基因组计划建立了迅速测定小部分天然蛋白质结构的可行性,但重要的瓶颈仍然是大部分天然蛋白质结构的迅速测定问题。例如,结构基因组计划集注在超嗜热蛋白质,但是他们在 *E. coli* 表达时,约 50% 的靶蛋白形成包涵体或不溶性聚集体。目前,解决这个瓶颈问题的策略可以分为两类[1]:①广泛检试策略,即应用天然可溶性蛋白质的条件进行广泛检试;②补偿策略,即应用修饰的蛋白质顺序,使其适于结构测定并保持它的天然构象。应用天然蛋白质顺序的优点是蛋白质很可能保持它的功能,缺点是在后翻译修饰过程中的晚期所进行的表达,再折叠、溶解(增溶)和结晶条件进行搜索、耗时、耗资、并且只有少量的靶蛋白。相反,应用修饰蛋白质的优点是适于高通量结构测定,顺序广泛搜索可以在早期阶段进行,检出的可能性更大。对工程可溶性蛋白质而言,依赖 GFP 体外定向进化技术可以简单地鉴定出高度可溶蛋白质的突变体,这个途径的潜在缺点是所产生的蛋白质有时可能是非功能的。

结构基因组学提出另一个有效的策略[2]:①由蛋白质族或根据重要生物化学功能选出靶顺序;②这些靶是错误折叠或不溶,并且不服从常规的稳定化方法,例如再折叠。但适于高通量结构测定,并可由定向进化和筛选得到。例如,采用 GFP 折叠报告的定向进化改进了核苷二磷酸激酶(NDP-K)的折叠和溶解性,进化的晶体含有 6 个点突变,有 NDP-K 酶活性。X 射线衍射解析结构的结果表明基本上与紧密相关的结构同源物相同。这个工作证实依赖 GFP 定向进化途径是有效的。

为了获得可溶的正确折叠蛋白质,常规途径包括低温表达,采用各种强度的启动子,各种提高溶解性的融合标签,修饰生长介质以及再折叠[3]。这些途径有时是成功的,但是对高通量应用来讲不是理想的。这是因为:①每个目标都要进行各种操作;②在实现目标之前条件试验是有限的;③这些方法不能用于修饰蛋白质的内在折叠稳定性和溶解性,甚至在最初表达或再折叠实验过程中以可溶形式获得的蛋白质组分,在随后的加工和浓缩时仍然可能发生不可逆的凝集,这是结构基因组学另一瓶颈问题。为了解决这个瓶颈问题,另一种策略是以某种方式改进目标蛋白质内在的折叠,产生稳定性和可溶性的蛋白质。当结构已知时,定位突变有时可以改进蛋白质的溶解性。*In silico* 方法也可以用于预测稳定的突变,但是这些突变对折叠途径的潜在有害效应是不能预计的。根据功能,有时可以设计出体内筛

选/选择方法，应用高通量液体处理技术，用细胞裂解液可以提高活性筛选[4]。由于它们需要目标蛋白质的结构或功能信息，所有上述方法都缺乏普遍性。对于大量各种目标蛋白质来讲，不是所有蛋白质的活性都容易检出。因此设计高通量功能检验可能是不实际的。这里介绍筛选/选择可溶蛋白质的一些方法，尤其是适合结构基因组学的一些方法。

最近发展的蛋白质可溶性筛选和选择方法基本上可分为 3 类：①融合报告法；②宿主细胞应激反应法；③直接检验方法。

8.2.1　融合报告

融合报告法(图 8.1)是一种检试蛋白质与易检验的报告蛋白质以遗传融合形式表达。检试蛋白质的折叠和可溶性是通过融合的报告结构域活性进行筛选/选择。

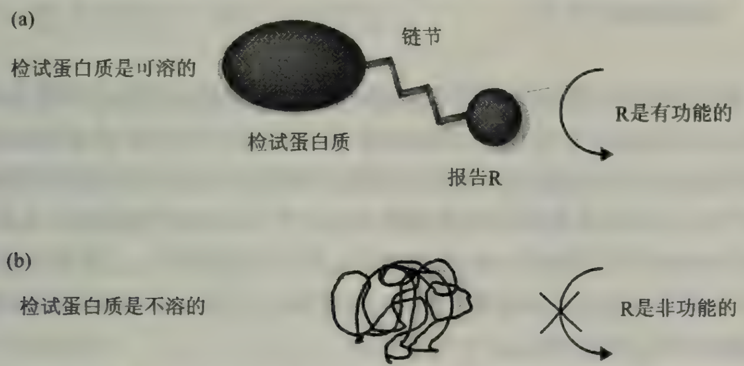


图 8.1　融合可溶性报告的原理

易测功能的报告结构域经柔性链节融合到检试结构域。(a)当检试蛋白质是可溶时，报告结构域是功能结构域，它的活性可以检测。(b)当检试蛋白质是不溶的或错误折叠时，报告结构域的功能处于功能和非功能之间。报告结构域可能部分具有易测功能。检试的方式只取决于报告的功能，其可能是可选择的标记蛋白质，例如酶或抗菌素蛋白质，或可筛选的标记，例如酶或荧光蛋白质。完整的蛋白质或蛋白质片段都可以用[11]

8.2.1.1　绿色荧光蛋白质(GFP)折叠报告法[5]

在该方法中检试的蛋白质的 C 端与 GFP 融合表达。GFP 的荧光产量传导融合蛋白质是否成功折叠的信息。表达折叠且可溶的融合蛋白质的细胞比缺乏折叠不溶的融合蛋白质的细胞荧光强度大。GFP 只是为了监控检试蛋白质的折叠产量。用这个方法可以工程蛋白质的可溶性[6~11]。

GFP 定向进化过程大体如下：每轮 GFP 定向进化是由两阶段构成[10]。第一阶

段是通过易错 PCR 和 DNA 改组建立 GFP 融合突变体库,第二阶段由突变体库中选择 *E. coli* 克隆,其与野生型相比,产生较强的荧光。GFP 融合到检试蛋白质 C 端作为上游蛋白质适当折叠的报告。从选择的克隆得到突变体基因,荧光提高的突变体的基因用于下一轮突变体库的制备。在三轮正向进化以后(没有同野生型回交),最终选择出具有最大荧光强度和溶解性改进的蛋白质。

8.2.1.2 氯霉素乙酰转移酶(CAT)融合法

Davidson 等[12]证实 *E. coli* 在含有氯霉素的介质中存活率,是与 N 端同 CAT 融合的检试蛋白质在细胞中可溶性表达呈正相关性,氯霉素提供了选择压力。Arnold 等[13]应用该法富集可溶性突变体库。

8.2.1.3 PROSIDE 技术

蛋白质可溶性往往高度依赖于稳定性和折叠。噬菌体展示技术能够选择稳定性提高的多肽。Shmid 等[14]应用 PROSIDE 技术选择小的冷休克蛋白质。他们证实两种选择压力(变性剂或提高温度)为蛋白质稳定化提供两种不同的策略。Baker 等[15]设计了一种选择策略以捕集熵降低的各种肽(增加折叠度)。在该方法中,各种肽展示在 SH2 结构域允许的部位,依次展示在噬菌体上。具有低熵(即稳定和闭合的 N 端和 C 端)的各种肽保留 SH2 结构域的功能。该噬菌体库在磷酸酪氨酸珠上选淘以回收折叠的肽。

8.2.1.4 质量控制

Barberis 等[16]发展了一种融合报告系统,他们称为质量控制(quality control)。采用可选择的遗传标记以鉴定可溶性单链抗体(scFv)。该系统在酵母中操作,采用了经典的双杂交系统与组成类似物融合。scFv 被融合到可选择的包含转录活化结构域(AD)和来自 Gal Ⅱ P 的肽片段标记蛋白质。如果 scFv 是稳定的,并且是可溶的话,融合的 AD-Gal Ⅱ P 结构域可以同结合的 DNA Gal4(1 ~ 100)片段结合,因此活化 *HIS3* 和 *LacZ* 报告基因的转录。在适当菌株背景和选择介质上,*HIS3* 的表达允许宿主细胞存活(适应选择),而 *LacZ* 允许蓝/白筛选。这样,表达可溶性 scFv 细胞将存活,而表达不稳定或不溶 scFv 的细胞将死亡,因为缺乏可检验标记 *HIS3* 的表达。在可溶性和细胞存活之间有很好的相关性。该系统是否适于结构基因组学正在探索中。

8.2.1.5 LacZ α(β半乳糖苷酶 α 肽)补充可溶性报告检验

LacZ α 肽由 100 个氨基酸组成。LacZ α 肽标签与检试蛋白质融合[17]。如果融合的蛋白质仍是可溶的,并且标签物不被掩蔽的话,那么,LacZ α 肽可以结合失活的 LacZΩ,恢复β半乳糖苷酶活性。分裂蛋白质的遗传补充具有以下优点:①小的标签可以预期对检试蛋白质的折叠和稳定性影响不大;②由于补充只取决于可溶形式中标签的可及性,肽标签可溶性报告对于监控缓慢蛋白质凝集过程可能是特别有效的;③LacZ α 系统可作为可选择的标签,应用最低介质(乳糖作为碳源)在适当宿主中存活。这可以选择改进的可溶性蛋白质大库($1 \times 10^{7\sim9}$)。LacZ α 可溶报告是否适于结构基因组学仍需探索。

8.2.2 宿主细胞应激反应报告

Bently 等[18]应用宿主细胞应激反应报告可溶性和折叠,这是非侵害的应激报告。它是根据热休克启动子驱动 GFP 表达以监控细胞对各种刺激的应激反应。例如热休克、渗透应激、乙醇和诱导。作者应用差式 mRNA 表达和 DNA 基因芯片检验 *E. coli* 细胞对各种应激因子的反应。当重组蛋白超表达时,热休克反应启动子 ftsH、clpP、ion、ompT、degP、groEL、aceA 和 ibpA 向上调节。应用驱动应激反应启动子报告基因,可以鉴定强烈影响体内蛋白质折叠产量的沉默突变[19]。Lesley 等[20]更严格地研究了 *E. coli* 启动子,当错误折叠蛋白质超表达时,其是最强并特异地被活化。含有 *E. coli* ORF 的排列的基因表达芯片用于评价 *E. coli* 超表达不溶蛋白质对可溶蛋白质的各种基因的表达。当错误折叠蛋白质表达时,某一基因特异地活化。根据这些实验,选择小热休克蛋白质 ibpA 的启动子控制 *LacZα* 报告基因的表达,它们可以区分可溶、部分可溶和不溶聚组氨酸标签的重组蛋白质。

8.2.3 直接检验

Knaust 和 Nordlund[21]应用滤板和斑点印迹发展了聚组氨酸标记的可溶性蛋白质表达的方便筛选方法。全细胞裂解液放在多孔滤板上,抽真空。可溶性组分通过滤膜,并在多孔板上收集。保留在滤板上的不溶部分应用变性剂溶解,并在第二个多孔滤板收集。膜上的组分洗涤,用抗聚组氨酸抗体探针并展开。用双盲实验也可以区分 96 孔板随机排列的两种蛋白质(一种是可溶的,另一种是不溶的)。该方法的缺点:可以筛选的库的大小不仅受有效的塑料器皿的限制,而且抗体斑点需要结合、洗涤和探针的冗长循环。

Peabody 等[22]发展了一种改变聚合和可溶性蛋白质的筛选方法。在营养琼脂平板上生长 *E. coli* 克隆表达质粒编码 MS2 病毒壳蛋白的突变体。该克隆铺上含有裂解试剂和抑制生长的抗生素琼脂层,可溶性蛋白质由克隆扩散到琼脂中。捕获蛋白质的膜,应用抗 MS2 壳蛋白抗体探针,并且用方便的 ELISA 技术检验。斑点的直径和强度与可溶性蛋白质的量以及蛋白质的扩散速率正相关。可溶性突变体的二聚体扩散的快,给出大的斑点。可溶性蛋白质同不溶性突变体相比给出更强的斑点。显然,该方法也采用一般标记物抗体(即聚组氨酸抗体)。目前,用该法 Petri 平板每 8cm 筛选出几百个克隆,但在克隆之间需要适当空间以避免个别克隆信号的重叠。

8.3 基因组改组

发酵占有 30~50 亿美元的世界市场。通过单细胞内的多酶催化可以将简单的原料转为高附加值产品。发酵产物的范围有高附加值天然产品、蛋白质药物、工业酶和化学品。科学家们不断的寻找催化路线,由可更新生物量产生商业的化学品,为此经常进行鉴定或构建菌株以改进化学转化。然而,第一代催化剂很少满足商业化执行的标准。限制生物催化剂应用的因素包括低产量、重现性、效价、过程条件的耐受力(例如极端温度、pH 或溶剂或溶质浓度)等,改进上述情况需花大力气。

用细胞将底物转为产物,最理想的是使细胞(即生物催化剂)能高密度迅速生长,能在最低生长情况仍保持高活力,能利用廉价的底物并以高速率分泌目的产物,很少或没有其他的副产物等。为了达到此目标,人们采用了各种方法。Patnaik[23]和 Zhang 等[24],证实了组合方法是改进细菌表型的有效方法。该方法涉及完整基因组改组——原生质体融合——少量的母体菌株表型有明显改进。在下述两种情况得到证实:①泰乐菌素(tylosin)生产的改进。泰乐菌素是一种弗氏链霉菌(*Streptomyces fradiae*)产生的复杂聚酮抗生素。②乳酸杆菌(*Lactobacillus*)耐酸菌株的分离,以使其适用于低 pH 发酵生产乳酸。在上述两种情况下,仅仅几轮基因组改组就可得到表型明显改进的菌株。

虽然随机突变和选择方法已成功地选出许多工业菌株,但是这些方法对于改进细胞的表型是困难的,因为表型是由分布于整个基因组的一批基因决定的,对其所知甚少,甚至不知。实验表明在组合途径中,细胞的随机突变或同源基因改组对于改进细胞表型是有效的。组合途径的成功取决于起始突变体的选择,遗传重组过程的效率和选择方法的灵敏度。Zhang 和 Patnaik 等证实原生质体融合允许基因组发生重组——基因组改组,基因组改组在原理上是改进主要的生产途径和表型,获得的菌株经济实用。基因组改组具有很大的潜力,特别是与其他方法组合。该技术的关键特征是通过改组表型改进的突变株的基因组而形成候选库,产生许多

改进的细胞。

假定"N"基因影响表型，例如产物途径，但我们不知道决定细胞最佳表型的基因数和确切的组合，这样，表型库可能包括细胞含有"N"基因所有可能的组合，即 2^{N+1}细胞库。当 $N=30$ 时，相当于十亿个细胞。另一重要考虑是最终需要表型最佳化。应当指出 Patnaik 等分离出的 *Lactobacillus* 菌株是耐酸（在 pH3.8 生长），但乳酸的产率不如野生型菌株。按逻辑来讲，下步应当进行基因组改组以改进耐酸和产率，但这样会产生两个表型选择的相容性问题。

图 8.2 说明实际基因型产生两个特殊表型相容[图 8.2(a)]或不相容[图 8.2(b)]，也就是说首先选择的菌株是一个表型最佳化的，然后进行另一个表型的最佳化或者所有表型同时最佳化。在表型抵触的情况下[图 8.2(b)]，将需要确定反映每个表型全部意义的特殊权重的组成表型，以指导整个过程最佳化。

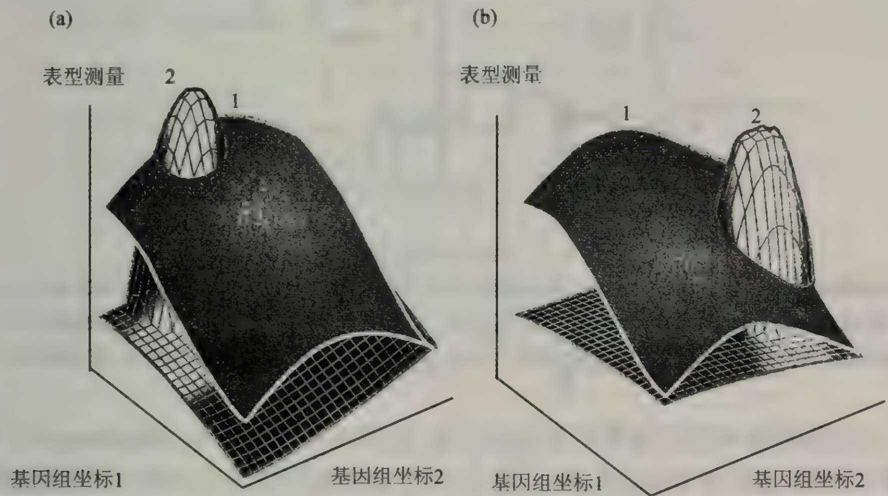


图 8.2　在多目标菌株选择中，表型与基因型关系图解

(a)在基因型空间同样范围内，表型 1 和 2 最大限度增加，这样可能获得满足两个标准的菌株。(b)与(a)相反，在这种情况下，需要选择或者两个表型之一或者反映每个表型相对重要的组成指标

通过基因组改组改进表型无疑是细胞改进的重要里程碑，对应用来讲，也是如此。

8.3.1　基因组改组改进泰乐菌素生产菌表型

进化是一种遗传变异和表型选择的连续过程。在选择的群体内的重组，新的突变体组合，扩增了群体的遗传多样性，从而改进了群体内个体的特性。虽然经典育种针对整个基因组，但它只允许每代两个亲本之间的重组。相反，DNA 改组针对 DNA 片段并允许每代多个亲本之间的重组，多亲本复合子代可以加速定向进

化。这样,经典育种和 DNA 改组的组合提供产生“复合子代”组合库的一种方法。Zhang 等[24]应用这种方法改进了完整微生物基因组。

微生物的定向进化传统上是通过经典菌株改进(CSI)的无性过程(连续随机突变和筛选)完成的,它指引商业微生物的开发。在 CSI 每循环过程中,由改进的突变体群体鉴定出单个最优良特性突变体。每个循环所产生的改进的突变体群体库包含有益遗传多样性,这个群体的改组应当产生最优良特性的突变体新的组合(图 8.3)。Zhang 等[24]选择改组链霉菌(*Streptomyces*),这是因为它们是商业抗生素生产菌株,有长期的 CSI 历史,有现成的有效重组方法。

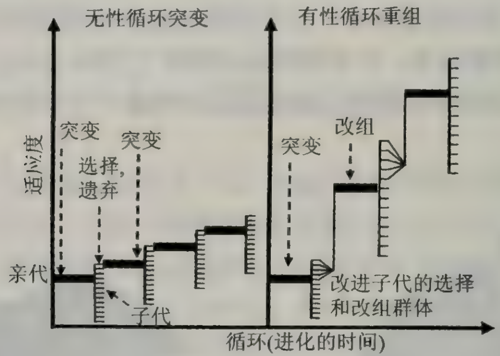


图 8.3　无性与有性进化比较

无性进化是累积个体突变的连续过程。最佳选择结果只获得单一突变体。它是缓慢的,只是群体内个体进化,信息不能共享,遗传多样性丧失,有害突变积累。有性进化允许群体内信息共享,选择群体内的匹配强化了有益突变的组合和有害突变的丧失。这样,有性进化产生的群体含有适应度大于亲代的个体

虽然在细菌细胞之间重组有各种方法,但原生质体融合在 *Streptomyces* 之间是最有效的,重组效率大于 0.20。“复合子代”分布的改进可以由混合原生质体群体的循环融合(recursive fusion)实现(http://www.nature.com)。每个库的融合是模拟 DNA 改组的每个循环反应,并且循环融合应当产生群体的有效改组。为了检验这种思想,通过循环原生质体融合改组了 *S. coelicolor* 菌株。在非选择条件下 4 个菌株的原生质体被混合、融合和再生。再生原生质体的孢子形成第一个融合库(F1)。同样 F1 群体再用于制备原生质体、融合和再生,这个过程重复 4 次结果形成 F4 群体。单一融合产生的双标记子代(10%)、三标记(0.4%)、四标记子代(0.000 07%)。然而循环融合,双一(60%),三一(17%),四标记子代(2.5%)在复合子代中增加 40 ~ 10^5 倍。这样循环原生质体融合成功地改组了 4 个天蓝色链霉菌(*S. coelicolor*)菌株,并且可以作为选择表型的群体。

弗氏链霉菌是用于生产泰乐菌素(tylosin,一种复杂的聚酮抗生素,用于治疗革兰氏阳性菌引起的感染和动物生长促进剂),泰乐菌素生物合成途径见图 8.4。

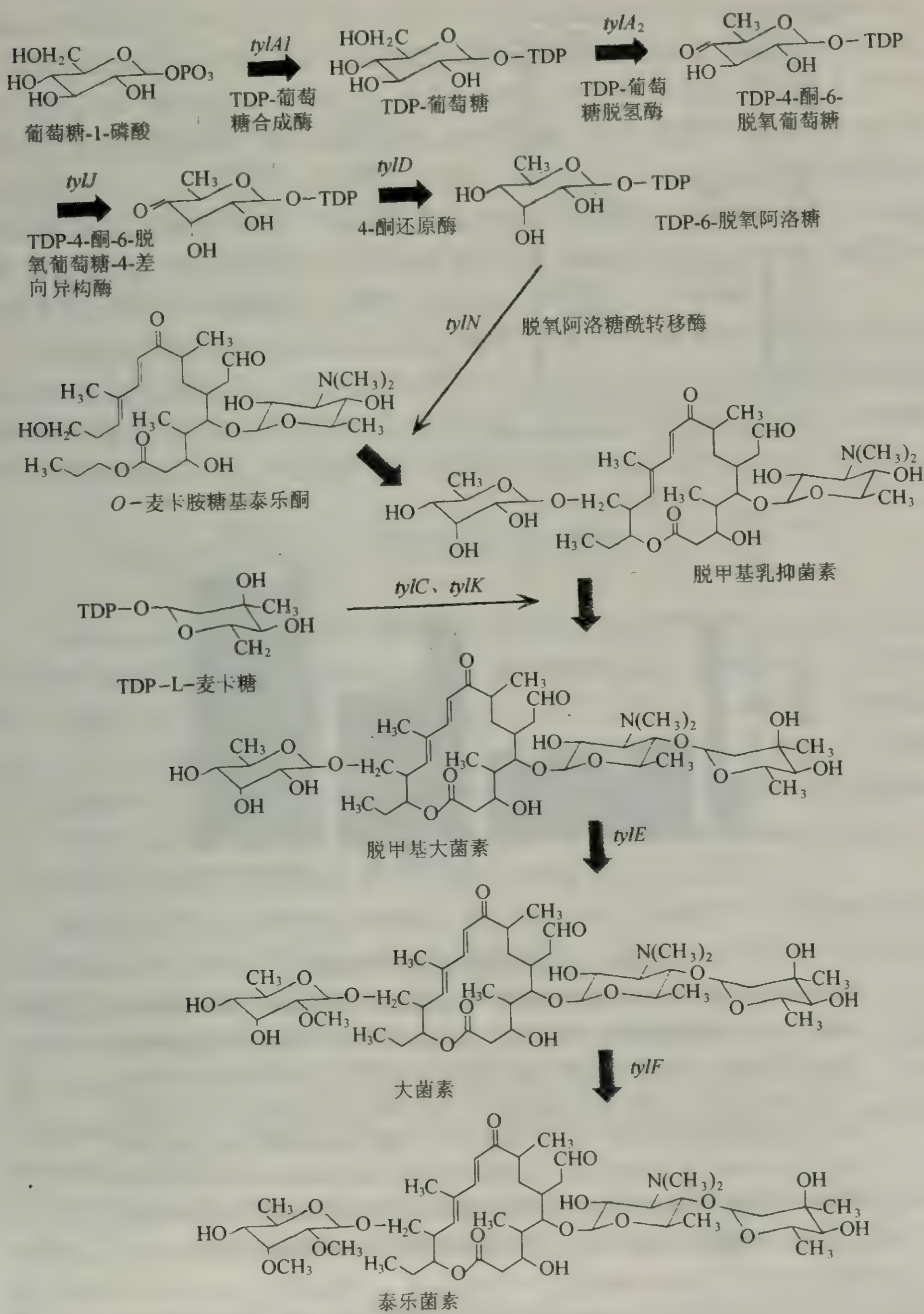


图 8.4　泰乐菌素生物合成途径[25]

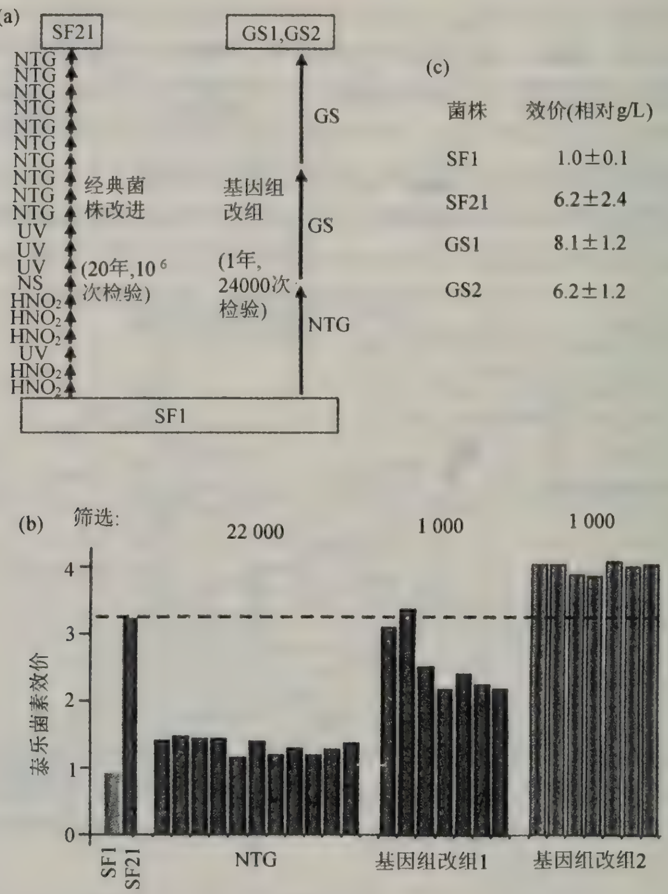


图 8.5　基因组改组与经典育种比较

(a) 弗氏链霉菌生产泰乐菌素,经典菌株与基因组改组菌株的比较。SF21 是 SF1 经 20 轮 CSI 突变和筛选得到的。GS1 和 GS2 是 SF1 经典改进菌株的群体通过两轮基因组改组得到的。经典的途径需要 20 年并近百万次筛选,而基因组改组只需要一年和 24 000 次筛选。图中:HNO_2,亚硝酸;UV,紫外辐射;NTG,亚硝基胍;NS,中性选择;GS,基因组改组。(b) SF1,SF21 在 300μl,96 孔发酵产生泰乐菌素(点线标明 SF21 产生水平);NTG,由 22000 筛选出 11NTG 改进的突变体;基因组改组 1,7 株是由 1000NTG 群体中筛选鉴定的;基因组改组 2,7 株是由 1000 基因组改组 1 筛选鉴定的。(c) SF1,SF21,GS1 和 GS2 经 250ml 摇瓶发酵产生的泰乐菌素(相对值)

菌株 SF1 是弗氏链霉菌(*S. fradiae*)天然分离得到的,而 SF21 是高效价产生泰乐菌素菌株(它是 SF1 经 20 个 CS1 循环得到的)[图 8.5(a)]。应用基因改组加速由 SF1 产生新的高产泰乐菌素的菌株。为了经基因组改组产生各种群体,应用亚硝

基胍(NTG)作为 SF1 诱变剂进行一轮 CSI,由 22 000 突变体中用 96 孔板筛选出 11 个菌株(以 SF1 泰乐菌素效价作为选择标准)进一步进化群体[图 8.5(b)]。制备原生质体,11 等份混合,循环融合。由改组的第一个循环筛选出 1000 子代,其中 7 株产生的泰乐菌素高于 NTG 诱变亲代,有的菌株产泰乐菌素的水平可与 SF21 相比[图 8.5(b)]。这 7 株进一步改组,筛选出 1000 子代。其中 7 株进一步改进泰乐菌素效价,在高通量条件下产泰乐菌素高于 SF21[图 8.5(b)],最好的两个改组菌株,GS1 和 GS2 同 SF1 和 SF21 在 250ml 摇瓶中比较,GS1 和 GS2 产泰乐菌素高于 SF1 9 倍,对于 SF21 在效价上无统计差别[图 8.5(c)]。

总之,两轮基因组改组足以完成以前需要 20 轮 CSI 所得到的结果,理论上,这是 CSI 的无性过程和基因组改组的超性过程之间的差别。但实践中,在选择的群体内循环重组可以明显增加菌株进化的速率。将改组技术由亚基因组片段扩展至整个细胞和代谢工程,不需要顺序信息或高级的遗传工具,简单易行且有效。

8.3.2　基因组改组改进乳酸生产菌表型

Lactobacilli 生产乳酸菌株。乳酸是长期使用的食品添加剂,最近也作为其他化学品的重要原料,例如聚乳酸(PLA)、乙醛、聚丙烯二醇、丙烯酸和 2′,3′-戊二醇等。发酵时 pH 对乳酸纯化过程最为重要。虽然生产时 *Lactobacillus* 发酵保持 pH5.0～5.5,但希望保持乳酸的 pKa 约 3.8 或以下。在这种低 pH 情况下,产物的实际比例是以自由酸形式存在,并且可以由发酵培养液中直接用有机溶剂提取纯化。在较高 pH 时,乳酸盐是主要形式,纯化耗时、耗资。改进微生物在低 pH 生长生产乳酸可以降低产品纯化成本,这是重要的商业目标。

通过插入突变所分离的乳酸球菌(*Lactococus lactis*)耐酸突变株的特性分析指出对抗低 pH 有 18 种不同位点和多重保护机制。*Lactobacillus* 酸适应的蛋白质分析鉴定出 63 种不同的蛋白质,它们是协调诱导产生的。同样,*E. coli* 中酸反应是几种调节剂诱导各种位点。不了解细菌的 pH 耐受机制,就无法工程 pH 耐受性菌株。此外,研究所用的 *Lactobacillus* 菌株是商业菌株,未作遗传学上鉴定。这样,工程该菌株缺乏更多的遗传手段。因此,基因组改组可能是改进未鉴定和复杂机体表型的理想策略。

目前,改进工业微生物的方法有经典菌株改进(CSI)的随机突变和合理代谢工程方法。虽然 CSI 是粗糙的,但时间和资源集中。合理途径是信息和手段集中,并且依靠模型(简化的生物系统)。分子育种应用到微生物育种,在泰乐菌素(tylosin)生产中已得到有效的证明。基因组改组也用于改进 *Lactobacillus* 菌株的耐酸性,具体操作如下。

8.3.2.1 起始群体的选择

应用两种经典方法由野生型 *Lactobacillus*(LB-WT)产生耐酸突变体群体。第一个群体(pop-adap)是通过恒化器调节 LB-WT 菌株适应低 pH 获得的(逐渐降低 pH 直到 pH4.1,该 pH 时严格地抑制 LB-WT 生产)。第二个群体是应用亚硝基胍(NTG)突变体库(pop-NTG)在 pH 梯度板上选择并富集野生型菌株的耐酸突变体。pop-adap 和 pop-NTG 同 LB-WT 厌气培养,并在 pH 梯度板进一步选择。由 pop-adap 和 pop-NTG 分离出亚群体 pop1 和 pop2(表 8.1),这些耐酸群体作为基因组改组的材料。

表 8.1 菌株的设计操作

菌 株	分离的个体菌株
LB-WT	Cargill Dow 乳酸产生菌
pop-adap	适应于低 pH 的 LB-WT 的群体
pop-NTG	LB-WT 的 NTG 突变体库
pop1	在 pH 梯度板上富集 pop-adap
pop2	在 pH 梯度板上富集 pop-NTG
F1	pop1 和 pop2 之间一轮融合之后产生的群体
F3	pop1 和 pop2 之间 3 轮融合之后产生的群体
F5	pop1 和 pop2 之间 5 轮融合之后产生的群体
F5-37、42、67、79、95	由 F5 库选择出的克隆进行摇瓶鉴定

8.3.2.2 基因组改组

基因组改组扩增了遗传多样性,新的群体代表遗传多样性组合库。改组两个 pH 耐受群体(pop1 和 pop2),经五轮以库的方式循环原生质体融合,每轮产生的原生质体样品保留分析,然后作为下一轮融合的原生质体资源,形成的群体应用酸梯度板筛选出改进的耐酸个体。新改组 *Lactobacilli* 无论在液体还是在固体介质中较之野生菌株更耐受低 pH。在 pH4.0 时,改组菌株产乳酸高于野生型 3 倍以上。这说明基因组改组可以广泛地用于改进工业微生物的表型。

8.4　生物合成途径定向进化

转基因作物正迅速的发展，这是因为它成本低、方便农民以及环境又好。2000 年美国有 25% 的玉米、54% 大豆和 61% 的棉花是转基因作物。第一代转基因，包括“输入品质”，减少数量和降低成本。含有苏云金芽孢杆菌（*Bacillus thuringiensis*）Cry 蛋白质（BT）的作物控制昆虫和减少杀虫剂的用量。BT 的作物的应用已经减少棉花和玉米生产中杀虫剂的用量。嘉磷塞（glyphosate）是一种环境友好的除莠剂，它对哺乳动物、鸟和昆虫无毒。抗 glyphosate 的大豆已经减少了其他除莠剂的用量。

转基因作物的下一代将包括附加的“输入品质”以控制植物疾病、昆虫和莠草；将来的研究将使作物增加产量，低施肥和抗寒，抗旱以及抗盐；也将效力于提高农产品的质量，提供新的工业原料，并且能产生治疗作用的植物。

近 20 年，科学家发展了改良作物质量的许多策略和方法。昆虫的控制，植物的抗寒，疾病控制和其他方面已经取得了积极的成果。但还存在一些问题，诸如基因尚未充分发挥功效、成本高等。有几种技术可以用于改良农产品，它们包括依据蛋白质结构的合理设计，改进天然酶活性的随机突变和筛选，特别是模拟自然过程的定向进化已为植物的改良做出了贡献[26]。

8.4.1　β-葡糖苷酸酶定向进化

β-葡糖苷酸酶（β-glucuronidase，GUS）广泛作为植物基因表达的报告。酶作为报告的一种限制是当组织用戊二醛或甲醛固定时其活性丧失。*E. coli* β-葡糖苷酸酶基因组改组（图 8.6）[27,28]，形成的突变体用底物 5-溴-4-氯-3-吲哚-β-D-葡糖苷酸（X-gluc）比色筛选。在第一轮改组中，9000 克隆在颜色展开之前用 0.2% 戊二醛处理 20min 后筛选。9 个克隆有活性，并超过母体。在第二轮，6000 克隆用 1.0% 戊二醛处理 20min，含有突变酶的 9 个克隆较之第一轮筛选出的有更大的活性。野生型母体用 0.44% 戊二醛处理 20min 后，酶活丧失 99.6%，但是第三轮 DNA 改组的大多数抗性酶用 0.2% 戊二醛处理仍保留 78.1% 酶活性。进化酶已经表明在用甲醛固定后浸染非洲爪蟾（xenopus）胚是有效的。

由于植物表达内源 β-半乳糖苷酶，*LacZ* 不能用作报告基因。取而代之是 *E. coli* β-葡糖苷酸基因（*gusA*，以前用 *uidA*）作为植物的报告基因，已广泛应用十年了。显色和荧光 GUS 底物已合成，允许非放射活性检验。在各种条件下 GUS 是稳定的和有活性的，甚至与其他顺序融合时也是如此。

然而，GUS 作为报告的有效性是受 3 种方式制约：①许多动物系统，某些植物

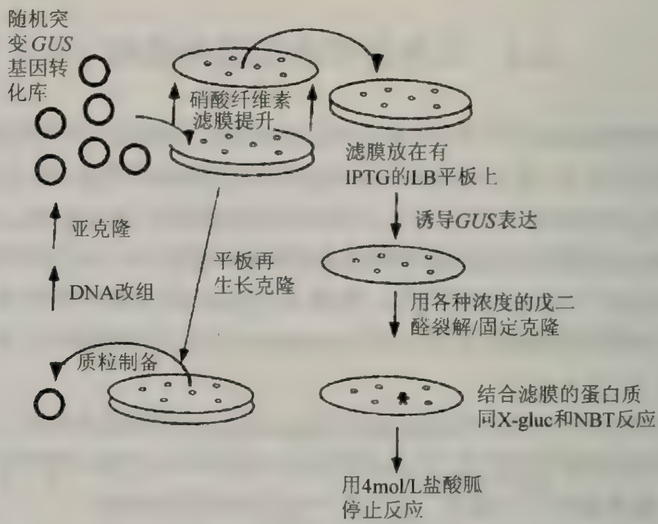


图 8.6　抗戊二醛β-葡糖苷酸酶(GUS)的筛选

随机突变β-葡糖苷酸基因(gusA)库亚克隆到可诱导的表达载体上并转化到*E. coli*中,形成的克隆转移到硝酸纤维滤膜,其铺在含有诱导剂的琼脂糖平板上,并在37℃温育10~24h,结合克隆的滤膜在含有戊二醛缓冲液温育,然后转移到含有β-葡糖苷酸酶,X-gluc和NBT的组织化学指示剂。在4mol/L盐酸胍中短暂温育阻止颜色展开。由原来的平板分离出保持GUS活性的克隆,并进行下一轮DNA改组

以及植物联合细菌表达内源葡糖苷酸酶活性;②当组织被戊二醛和甲醛固定时,GUS活性明显降低,对组织结构的保存来讲,保持活性的交替换位是必要的;③上述考虑严重地限制 *GUS* 用作脊椎动物系统的报告基因。

醛失活酶主要是由于同表面可及的赖氨酸残基形成 Schiff 碱。当通过定位突变去除赖氨酸残基可以提供抗固定剂的突变酶。但许多表面赖氨酸是功能的关键并且不易改变。*E. coli*、人、鼠、老鼠和狗的同源物的顺序是已知的。在 *E. coli* 蛋白质中的27个赖氨酸残基有6个是其他种类共有的,这可能是主要的。然而,为了发现27个赖氨酸残基的组合可能被改变,为了增加对固定剂的抗性而酶没有失活,这样需要构建和检验大量的突变酶,因此为了改变GUS的表面化学,或者避免或者容纳醛修饰而没有丧失酶活性,采用了随机突变的途径来改变酶的底物特异性或热稳定性。在3轮突变、DNA改组和筛选之后,分离出一个有8个氨基酸改变的突变体,令人惊奇的是,已知戊二醛是修饰和交联自由胺,但只有一个赖氨酸残基突变。一般地,氨基酸改变发生在近于保守的赖氨酸,暗示酶的表面化学的选择,或者接受或者避免戊二醛修饰。实验结果表明可能有多种路线,突变体可能适应于固定或溶剂系统,最重要的是引入氨基酸使酶具有新的表面结合功能。

8.4.2　三嗪水解酶类定向进化

三嗪代表一类杀虫剂，其广泛用于农业。水解三嗪的酶能够由土壤和地下水中去除这些化合物。*Pseudomonas* 的 *atzA* 基因表达产物水解三嗪相关化合物 R1 位的氯和氟、不水解甲氧基、硫代甲基、氰基、叠氮基和氨基取代基[29]。TriA，即三聚氰胺脱氨酶(melamine deaminase)也来自假单胞菌(*Pseudomonas*)，其与 atzA 有 9 个氨基酸是不同的，TriA 和 atzA 用作 DNA 改组实验的母体以探索突变体作用的底物范围。1600 突变体库对着 50 个不同的三嗪化合物进行筛选，包括母体酶最适底物，不适底物和非利用底物。母体酶对其中 7 个三嗪化合物活性很低，改组酶库产生多个重组酶，较之最好的母体酶有更大的转化速度，改组库也含有水解不被母体酶利用的 5 个底物。这样，两个高度相关的天然基因的 DNA 改组产生大量的各种各样的酶，其中能够降解不同但相关的三嗪化合物。这说明 DNA 改组是获得降解土壤和地下水中杀虫剂污染物的新酶，也是植物中产生抗杀虫剂基因的有效途径。

苏云金芽孢杆菌(*Bacillus thuringiensis*) *Cry* 蛋白质成为成功转基因抗昆虫耐药性策略的基础，目前广泛用于农业。BT 作为转基因杀虫剂的定向进化是有潜力的。第一，受 BT 影响的昆虫谱相当窄。定向进化能够将其拓宽，这样减少化学杀虫剂在转基因植物的用量；第二，在转基因植物中，BT 蛋白难于高浓度表达产生农学上有关的表型；第三，定向进化可以有效地增加 BT 的特殊活性。

目前 BT(Cry1AB 和 Cry1AC)用于转基因作物，有效地控制几种棉铃虫属(*Heliothis*)害虫，而在 *Spodoptera* 对照中是无效的。在单一基因改组实验中，Cry1CA 改组并对着甜菜夜蛾(*Spodoptera exigua*)活性筛选。应用三层阶式筛选：第一层采用高剂量 BT，并且克隆是以活性或无活性划线；第二层采用一系列稀释 BT。在每个样品中蛋白质的含量非标准化，并且对每个活性克隆测定改进的速率。第三层筛选，一系列稀释重复 3 次，提取物的蛋白质含量标准化，测定昆虫杀死一半所需要的蛋白质浓度(即 EC_{50})。由大约 4000 基因突变体中筛选出 BT 突变体，其 EC_{50} 超过母体 3.8 倍。在另一基因改组实验中，基因突变体对 *Spodoptera* 活性 BT 是 Cry1CA 母体的 6.7 倍。其中最有趣的是几个突变体对玉米夜蛾、美洲棉铃虫(*H. Zea*)和 *Spodoptera* 都有活性。这些突变体中最有意义的是在控制 *H. Zea* 时像 Cry1AB 一样有效，并且较之 Cry1AC 在控制 *Spodoptera* 上改进 2 倍以上。上述工作清楚地证明，通过 BT 控制昆虫谱可以用 DNA 改组拓宽，这种技术代表了控制昆虫下一代转基因植物的新途径[30]。

定向进化允许分子生物学家模拟动物和植物育种操作来改进基因和途径，定向进化与植物和动物育种有相似但也有很大的差别。在常规育种中，交互局限于

两个母体,而 DNA 改组允许同时几个母体重组;常规育种涉及改进的植物和动物不敏感选择,而 DNA 改组允许改进的蛋白质多维选择。常规育种取决于植物或动物在改进条件下的生命周期,需几个月或几年,而 DNA 改组和筛选每个周期可以几周内完成。

8.4.3 类胡萝卜素生物合成途径定向进化[31~34]

类胡萝卜素(carotenoid)是植物和微生物正常发生的色素。它们的主要功能是通过猝熄光敏剂抗氧化损伤、同单态氧相互作用、并清除过氧自由基,这样阻止有害氧的积累。这种保护过程取决于它们的化学结构(多烯生色团的长度,端基的性质和它们含有的各种取代成分)。类胡萝卜素精确的化学结构决定它们的生物学性质,决定它们如何同其他的分子相互作用并整合到膜上。

过去 20 年,所积累的证据表明类胡萝卜素在防治心血管病和癌起重要作用。例如诱导皮肤癌的老鼠和鼠喂饲或注射 β-胡萝卜素 canthaxanthin(角黄素)能延缓肿瘤的生长。在其他动物研究中,已观察到能抗其他类型的癌(例如唾腺癌、结肠癌或乳腺癌)。至今已鉴定的 600 种以上天然发生的类胡萝卜素,其中大多数仅以微量生物合成中间体存在,目前研究方面的主要限制是各种各样结构的类胡萝卜素的供应问题。

类胡萝卜素的工业应用包括它们作为营养添加剂,医药、食品着色剂和动物饲料。类胡萝卜素的实际销售额大约 50 亿美元,并且市场正在扩大。至今,用于工业的大部分类胡萝卜素是化学合成的,例如虾青素(astaxanthin)、角黄素和 β-胡萝卜素,也有天然提取物。此外,β-胡萝卜素和虾青素也由微生物发酵生产,但市场份额不大。然而,作为更复杂的显露出药物潜力的类胡萝卜素,应用天然或一般修饰微生物发酵生产将是一种趋势。

最近分子生物学的进展已经提供了类胡萝卜素生物合成各种基因。不同机体基因组合按着途径的不同分支可以合成新的化合物。然而,这种工作刚刚启动。通过基因组合产生类胡萝卜素,tetracenomicine 及相关抗生素(经由聚酮途径)可能是一种有效途径。

应用基因重组技术已在非胡萝卜素细菌和酵母成功地合成类胡萝卜素。细菌、藻类、真菌和高等植物的胡萝卜素生成基因或 cDNA,大部分都能在 *E. coli* 中表达。各种类胡萝卜素的结构和它们的形成途径示如图 8.7。

天然产物包含各种各样化学结构和生物功能,许多有潜在应用价值的化合物在天然资源中仅以微量存在,并且难以或不可能化学合成,但可能用生物合成实现。生物合成途径的酶可以组合到天然和新代谢产物合成的新途径中。然而,依靠自然界存在的催化功能,不受限制的在工程有机体中合成所需化合物是不可

行的。

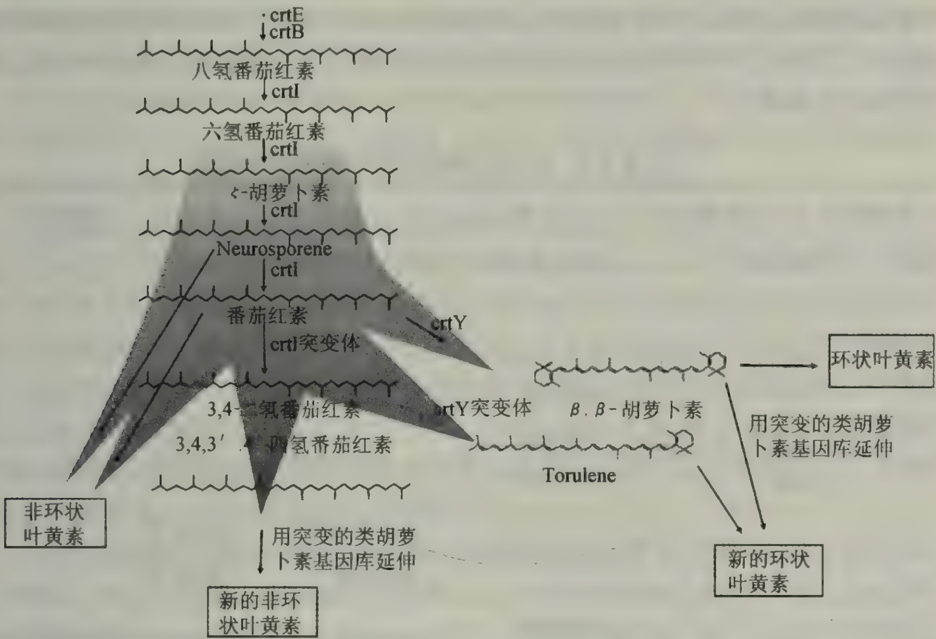


图 8.7　类胡萝卜素生物合成途径的定向进化

类胡萝卜素生物合成通过引入两个生物合成基因(*crtE*,栊牛儿基-栊牛儿基二磷酸合酶基因和 *crtB*,八氢番茄红素合酶基因)被工程到 *E. coli* 中,首先合成八氢番茄红素。该途径用改组的八氢番茄红素脱饱和酶(crtI)基因库延伸,在 *E. coli* 中产生完全共轭的四氢番茄红素的新途径。用附加的进化生物合成基因延伸这个途径可以产生各种各样新的环状和非环状类胡萝卜素结构,例如在 *E. coli* 中通过突变体环化酶(crtY)产生新的环状 torulene

分子育种的原理可以迅速增加天然产物的多样性,并且在简单的机体,例如 *E. coli* 中实现。为了建立新的生物合成途径,可以混合和匹配不同来源的基因,甚至不相关的代谢路线,在缺乏酶的结构或催化机制详细信息情况下,通过随机诱变,重组和选择产生新的生物合成功能。

类胡萝卜素生物合成证实了分子育种概念的正确性。通过“育种”生物合成基因和进化新酶功能,可以产生各种各样类胡萝卜素。在非类胡萝卜素生成的微生物(例如 *E. coli*)中生物合成类胡萝卜素需要延伸一般的类萜途径,该途径具有栊牛儿基(香叶基)-栊牛基二磷酸(GGDP)合酶(crtB)和八氢番茄红素合酶(crtE),产生 C_{40}类胡萝卜素八氢番茄红素(图 8.7),继之用八氢番茄红素脱饱和酶(crtI)去饱和,并进一步用(例如环化酶,羟化酶和酮基酶)催化修饰,结果产生各种类胡萝卜素。至今,用基因重组生物合成类胡萝卜素大多数基因是来自红细菌(*Rhodobacter*)或十字花科植物软腐细菌(欧文氏菌属)(*Erwinia*)。

为了能够在 *E. coli* 中合成新的类胡萝卜素，对八氢番茄红素脱饱和酶（crtI）和番茄红素环化酶（crtY）定向进化。这些酶定位在类胡萝卜素生物合成途径的重要分支点，并决定产生非环状或环状类胡萝卜素（图 8.7），类胡萝卜素生物合成的某些酶列入表 8.2。

表 8.2 类胡萝卜素生物合成酶类

基因型	来源	酶	底物	产物	所需辅因子
CrtE	*Eruinia*	GGPP 合成酶	FPP	GGPP	
			GPP	GGPP	
CrtB	*Eruinia*, *A. aurantiacum*	八氢番茄红素合成酶	GGPP		
CrtI	*Eruinia*, *A. aurantiacum*	八氢番茄红素脱氢酶	八氢番茄红素	番茄红素	FAD
			ζ-胡萝卜素	番茄红素	
			链孢红素	番茄红素	
CrtY	*Eruinia*	番茄红素环化酶	番茄红素	β-胡萝卜素	$NADH_2$
			γ-胡萝卜素	β-胡萝卜素	

八氢番茄红素脱饱和酶，或者在八氢番茄红素中引入两个去饱和产生 ζ-胡萝卜素，像在植物和蓝细菌一样；3 个去饱和产生链孢红素（neurosporene），像在红细菌（*Rhodobacter*）一样；或 4 个去饱和产生番茄红素，像在 *Erwinia* 和其他的光合成细菌一样。粗糙链孢霉的脱饱和酶在八氢番茄红素引入 5 个双键以合成 3,4-二氢番茄红素。脱饱和酶能够把 6 个双键引入八氢番茄红素，产生完全共轭的类胡萝卜素 3,4,3′,4′-四氢番茄红素，噬夏孢欧文氏菌（*Erwinia uredovora*）的八氢番茄红素脱饱和酶在一定条件下合成小量的 3,4,3′,4′-四氢番茄红素。第一个目标是通过进化一种有效的 6 步脱饱和酶来延伸去饱和途径以产生 3,4,3′,4′-四氢番茄红素作为 *E. coli* 中主要的类胡萝卜素（图 8.7）。

为了在 *E. coli* 生物合成途径中产生突变酶库需要两个质粒共转化：产生类胡萝卜素前体（作为靶酶的底物）的基因克隆在质粒上；进行定向进化酶的基因被克隆在另一质粒上。所有的酶都是在 Lac 启动子控制下表达，继之最佳化 SD 序列。欧文氏菌（*E. herbicola*）和 *E. uredovora* 的基因通过体外 DNA 改组（DNA shuffling）所产生的脱饱和酶库转化到合成八氢番茄红素 *E. coli* SM101（携带有 *crtE* 和 *crtB* 的质粒），结果产生 I14 脱饱和酶。

从红细菌(*Rhodobacter*)的链孢红素或其他光合成细菌中的番茄红素出发已合成各种各样的非环状类胡萝卜素,进一步脱饱和,羟基化和甲基化。还有其他细菌,例如 *Erwinia*,由番茄红素合成环状类胡萝卜素。这些修饰表明高度的混杂性,允许它们相等地作用在工程途径中的链孢红素和番茄红素。各种酶或它们的进化突变体可以进一步延伸生色团,导致在 *E. coli* 产生新的类胡萝卜素。下面将讨论通过定向进化环化酶产生环状类胡萝卜素生物合成的新途径。

细菌番茄红素环化酶通常把 β-紫罗酮环引在番茄红素的两端产生 β,β-胡萝卜素(图 8.7)。然而,在工程途径中,当链孢红素是由 *Rhodobacter* 脱饱和酶三步脱饱和产生或 ζ-胡萝卜素是由蓝细菌(蓝藻)(*Synechoccus* sp.)脱饱和酶二步脱饱和产生时,环化酶不仅能够环化 Ψ 端基到 β 端基,而且也能够环化 7,8-二氢端基(图 8.7 中类胡萝卜素结构)。各个单环中间体的合成证实酶分别作用于两端,环化反应机制只涉及双键 C1-C2(C1′-C2′)和 C5-C6(C5′-C6′)。因此在用 I14 的进化延伸脱饱和途径中,可以通过 DNA 改组产生番茄红素环化酶,其有效地环化 3,4-二氢番茄红素(3,4,3′,4′-四氢番茄红素前体)。

通过把基因组装在代谢途径和进化关键酶的分子育种是在 *E. coli* 合成新化合物的一种有效策略。定向进化允许我们工程两个酶的催化性质。用追加生物合成的基因并进一步进化,可以在 *E. coli* 产生更新的类胡萝卜素。由于许多新的生物合成途径可以在单一分子育种实验中产生,形成的机体代表新生物活性化合物的天然物库。最近开发的“金稻”(golden rice)含 β-胡萝卜素[34]。这些实验指出应用定向进化可以提高类胡萝卜素生物合成途径中特殊酶的活性并提高金稻中 β-胡萝卜素的水平。

8.4.4　L-甲硫氨酸生物合成途径定向进化

精细化学品工业正日益转向生物合成路线,以满足对映体纯药物迅速增长的需要。但生物催化剂与其他技术竞争尚有局限性,甚至手性产品的稍许修饰就得研究新的生物催化剂。

各种 D-和 L-氨基酸的工业生产的乙内酰脲过程已经建立,然而这个过程不能用于产生某些 L-氨基酸,例如 L-甲硫氨酸(L-Met),因为乙内酰脲酶对外消旋乙内酰脲底物转化表现出对映体选择性。该酶优选 D-5-(2-甲基硫代乙基)乙内酰脲(D-MTEH)导致 D-N-羰基甲硫氨酸(D-C-Met)的积累(图 5.24)由于 D-C-Met 转化为 L-Met 等,从而降低了 L-Met 生产能力,事实上,所有已知的乙内酰脲酶在这个反应中都表现出 D-选择性。

特殊的工业要求的酶可以通过突变/重组以及选择或筛选的重复过程产生。进化的脂肪酶,对 2-甲基癸酸酯对映体选择性水解,最好突变体产生(*S*)-酸,ee 值

>90%(野生型 2%)。这就是说,定向进化可以改进酶的天然对映体选择性。然而,对映体选择性如何转化是不清楚的。D-选择性酶转化为 L-选择性酶可能需要大的结构改变,应用定向进化可能达不到。

Arnold 等[35]发展了 *Arthrobacter* sp. DSM9771 乙内酰脲酶转化对映体选择性,用保持或增加全细胞催化活性生产 L-甲硫氨酸。这表明进化的乙内酰脲酶通过 *E. coli* 的三一酶途径改进 L-Met 生产。这是通过全细胞催化剂中关键酶的转化对映体选择性和增加总活性来实现的。所有已知乙内酰脲酶对 D-5-(2-甲基硫代乙基)乙内酰脲(D-MTEH)的选择性超过 L-对映体,结果导致 L-氨基酸中间体的积累并降低生产能力。应用随机突变、饱和突变以及筛选,使 *Arthrobaeter* sp. DSM9771 的 D-选择性乙内酰脲酶转为 L-选择酶,并且总活性增加 5 倍。全 *E. coli* 细胞表达的进化 L-乙内酰脲酶,L-N-甲氨酰化酶和乙内酰脲外消旋酶在不到 2h 由 100mmol/L D.L-MTEH 产生 91mmol/L L-Met。改进的乙内酰脲酶生产能力提高 5 倍。不需要中间体 D-甲氨酰化,甲硫氨酸的积累降低 4 倍(与野生型细胞途径比较)。高度 D-选择的乙内酰脲酶突变体也已发现。对映体选择酶通过定向进化迅速最佳化并引入多酶途径,可以产生改进的全细胞催化剂,可有效地生产手性化合物。

代谢途径产生各种各样的化学结构和生物活性化合物。人们估计仅植物就产生成千上万种相对低分子质量的化合物,鉴定的约 50 000 种。因此天然产物是发现新的有效化合物的巨大宝库。

为了增加天然产物的多样性,异源途径已被工程到重组机体中,例如生产小分子的工程 *E. coli*。在组合生物合成途径中,往往可观察到次级代谢各种酶的底物特异性松动,产生额外的天然化合物。该途径对新聚酮化合物合成是相当有效的,这是由于模块聚酮合酶固有的生物合成灵活性所致。然而,对于许多代谢途径来讲,代谢基因的组合组装却没有产生功能途径。

代谢途径的多样性指出了代谢基因高度进化的潜力。事实上,增加基因组顺序和结构以及提高初级和次级代谢酶活性是有效的方法,主要是通过各种途径中酶基因的复制来进化。例如,TIM 桶[$(\beta/\alpha)_8$ 桶]折叠是催化完全不同反应的许多酶族中所发现的最丰富的型主,而且它们有一个共同的祖先。次级代谢(包括降解途径)的各种酶,不像初级代谢的大多数酶,由于底物特异性的松动,往往在催化上是混杂的,并且在某种情况下可以产生一个以上的产物。次级代谢的可塑性是品质的选择结果,增加和保持化学多样性,而不增加适应度成本。定向进化策略与代谢工程一起可以模拟天然代谢进化创造新的途径,以产生天然和非天然分子的多样性以及异种生物化合物的降解。

8.5 病毒定向进化

遗传重组是许多病毒进化的主要驱动力，迅速改组的遗传信息导致具有新基因和新表型重组子的产生，体外 DNA 改组过程[36]（分子育种）加速了这种机制。多重同源母体顺序平行重组可以产生复杂重组子的多样库，由其中选择出所希望的改进病毒。

改组 6 种鼠白细胞病毒(MLV)的包膜顺序，继之选择产生具有完全新的向性的嵌合中国田鼠卵巢(CHOK1)细胞克隆。所选择的克隆的组成和性质指出母体顺序的这种特殊排列不容易通过天然逆转录病毒重组完成。这个例子证实分子育种在遗传上可以提高逆转录病毒高度进化的潜力，以获得希望的表型。当信息有限，多重复杂功能必须同时平衡时，分子育种可能是基因治疗、疫苗和病毒最佳化的一种有效工具。

DNA 改组可以有效的变更多个母体的功能顺序以改进表型。通过体外 DNA 族改组 6 种不同的母体 MLV 包膜顺序产生改组库。将这个库克隆到感染的 Moloney 主链（骨架）上产生 1×10^6 改组包膜克隆库[图 8.8(a)]。母体病毒(Moloney、292E、Friend2、Friend7、Friend9 和 Friend21)完全不感染 CHOK1 细胞，对其他细胞类型有各种感染性。从转染 293/G1 细胞库中制备库化的质粒 DNA。

通过细胞传代改组 CHOK1 和 Lec8 细胞混合库，选择出 CHOK1 感染性细胞[图 8.8(b)]。Lec8 细胞是 CHOK1 衍生的突变体，糖苷化途径有缺陷。它们的亲嗜性受体是更易接近，允许通过某种亲嗜性 MLV 感染，由转染的 293/G1 细胞产生的 Friend2、Friend9 和 Moloneg MLV 是能够相对有效地感染 Lec8 细胞。Lec8 细胞作为"桥"或中间体细胞，扩增和富集嵌合病毒，获得更低的感染活性的 CHOK1 细胞。这样的病毒完成生长优势，由于大量细胞对感染有效，然后选择，作为天然逆转录病毒进化的基线。

由感染的共培养细胞的基因组 DNA 所克隆的原病毒包膜顺序确定选择出现的优势种。没有从母体传代细胞回收到原病毒顺序。优势包膜顺序是一个嵌合体，其由 4 个片段组成。嵌合体包膜顺序不同于已发表的 MLV 相关株(已知感染 CHOK1)。虽然 3′LTR 也被改组，但该区域改变对 CHOK1 感染性是不需要的。

嵌合病毒的向性是不同于任何母体病毒(图 8.9)。通过 Lec8/G1 细胞传代，获得了对 CHOK1 的感染，而且它能效地再感染新的 Lec8 细胞。这与母体 Friend2、Friend9 和 Moloney 病毒相反。虽然这些可以有效地感染由转染 293/G1 细胞产生的 Lec8 细胞，但当通过 Lec8/G1 传代时它们的子代却降低了再感染新 Lec8 的能力，并对 CHOK1 细胞无感染性。当传代的母体病毒与 CHOK1/G1 和 Lec8/G1 共培养时，它们能迅速地被稀释掉。这样，当它们通过 Lec8 传代时，嵌合病毒和母体
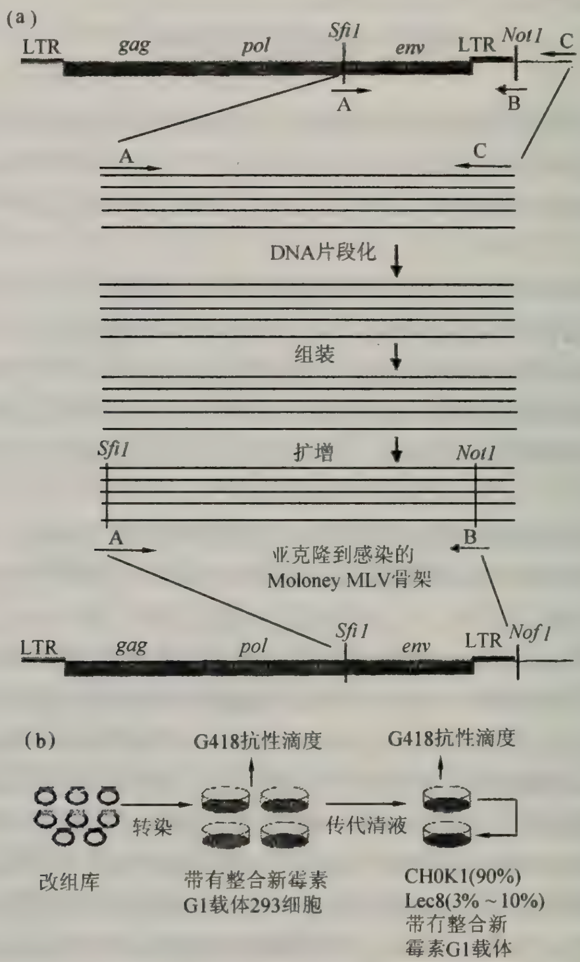


图 8.8 改组 MLV 包膜库的构建

(a) 病毒顺序的改组;(b) 传代库和 CHOK1 向性病毒的选择

病毒的感染性是不同的。

虽然通过亲嗜性 MLV 封闭 CHOK1 感染可以通过人工抑制亲嗜性受体的糖苷化而解除,但根本不明显。通过病毒包膜的设计操作来完成糖苷化的母体包膜不是有效的。通过 Lec8 传代的母体病毒不仅不能给予 CHOK1 感染性,而且也导致 Lec8 细胞感染性的丧失。通过改组产生新的嵌合包膜经历糖苷化的外遗传(epigenetic)。为了提高 CHOK1 细胞的感染性,两者需要给予新包膜的有利构象,只有体内逆转录病毒重组而没有体外改组不可能产生嵌合包膜。

通过改组相关天然病毒的顺序多样性可以构建高质量库,可以容易选择出性

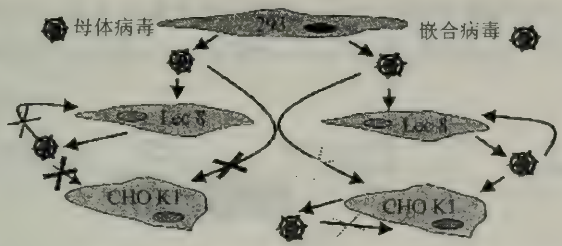


图 8.9　母体和嵌合病毒的不同向性

由 293 产生的嵌合病毒可以感染 Lec8,但不能有效地感染 CHOK1。当通过 Lec8 传代时,它们对 CHOK1 的感染性提高约 100 倍。它们也可以有效地再感染新的 Lec8 细胞。相反,当通过 Lec8 传代时,母体病毒降低了再感染新 Lec8 的能力,它们不感染 CHOK1,不管它们如何传代

质改进的组合病毒。分子育种可以扩展顺序空间并有效地搜索,而不需要更重要的信息。预期分子育种将成为病毒工程的重要工具。

用于人的基因治疗的逆转录病毒应是高纯和高效价的制剂。当用常规细胞培养系统生产逆转录病毒时,下游过程必须将产品浓缩至 100 倍。对生产而言,逆转录病毒载体的纯化和浓缩是至关重要的,对浓缩的物理过程非常敏感,结果不能有效地生产高效价的逆转录病毒载体。在浓缩(例如离心)过程中逆转录病毒载体的不稳定性似乎与病毒包膜蛋白质的性质相关。由于用水疱性口炎病毒 G 蛋白(VSV-G)取代逆转录包膜蛋白或“假型”允许超离心浓缩逆转录酶载体,然而,G 蛋白质的细胞毒性使这样的稳定包装的细胞系假型载体的生产在技术上是困难的,并且也拓宽了载体的宿主范围,这可能对于某些应用来讲是不利的。

鼠型 C 逆转录病毒的包膜蛋白质是由 SU、gp^{70}和 TM、p^{15E}组成的,其分别调节与细胞受体的结合和同细胞膜融合。研究指出在 SU 区释放后所发生的构象变化引发膜融合。SU 和 TM 亚基通过不稳定的二硫键连接。在纯化过程中,逆转录病毒的不稳定性被认为是由于 SU 结构域的丧失。以前已经指出在超离心的逆转录病毒的制剂中 SU 量减少。通过超滤方法纯化病毒的报道指出感染性丧失,而在高压下加工没有逆转录酶活性,这表明病毒核心仍然完整,而丧失了对感染所必需的表面蛋白质。

应用 DNA 改组可以加速重组子产物按所要求性质进化。以前 Soong 等[35]报道了通过 DNA 改组进化 MLV 的新向性。继之,Otto 等[37]成功地应用 DNA 改组改进了逆转录病毒的稳定性和加工产率。他们对 6 种亲嗜性 MLV 株进行 DNA 改组(图 8.10)。被改组的包膜区产生 5×10^6RCR 具有复制能力的逆转录病毒重组子库。这个库进行浓缩,在每个循环之后扩增幸存的病毒,分离出稳定性明显改进的病毒克隆。最好的克隆表现出降低母体病毒效价(降低 30 ~ 100 倍)的条件下效

价不损失。这些抗性病毒的包膜在 DNA 和蛋白质顺序上是不同的，都是来自多个母体的复杂嵌合体。这些研究表明 DNA 改组是改进基因治疗载体的有效工具。

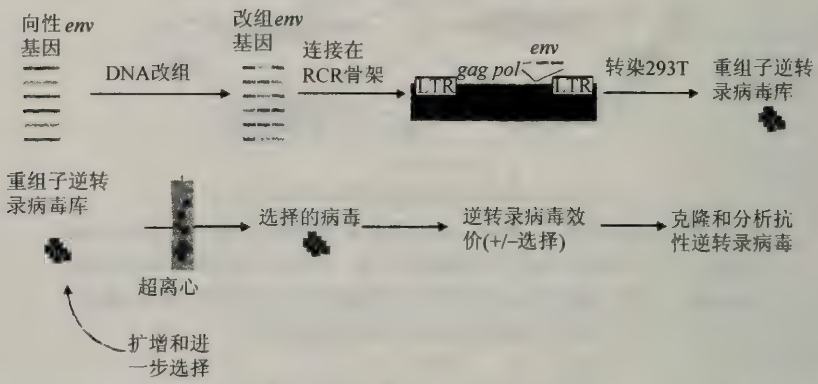


图 8.10　MLV 包膜基因改组实验图解

改组 6 种亲嗜性鼠逆转录病毒的包膜基因，并连接在 RCR 骨架上。通过瞬间转染产生重组子逆转录病毒库，在浓缩 VSV-G 假型载体所用的条件下离心。多轮离心的幸存病毒被克隆，以便进一步分析

8.6　微生物细胞定向进化[38]

经过漫长的进化，自然界创造出大量的结构不同和复杂的生物大分子，例如，各种抗生素、多糖、蛋白质、酶、核酸和其他的天然产物。许多次级代谢产物往往是由各种构件(building block)组合成模块结构(modular structure)，通过模块组装大大增加了生物分子的多样性。由于基因操作、基因组学、蛋白质组学、代谢工程和定向进化，工程微生物朝着新的非天然化合物生产方向发展。目前正试图生产新的生物化学试剂，特别是集中在抗生素、类胡萝卜素和多羟基链烷酸(表 8.3)。

微生物是各种化合物库源，也是能够生产化学合成难以或不能合成的各种化合物的工厂。微生物细胞定向进化包括组合各种途径的基因的组合；各种多结构域的酶衍生新的模块组合酶；定向进化途径和底物引导新构件途径，见图 8.11 所示。

(1) 新基因组合

根据生物合成途径的起源可以把机体分为两类。第一类是天然存在(或同源)生物合成途径的机体。同源途径可以用于开发新化合物的起始材料。例如，聚酮天然生产菌，链球菌属菌株已用来开发新的聚酮抗生素。第二类是工程(或异源)生物合成途径的机体。异源途径可以通过所要求的基因共表达来建立。这些基因

表 8.3　微生物细胞定向进化产生新化合物

修　饰	方　法	文献
多羟基链烷酸酯(或盐)		
含硫多羟基链烷酸酯(或盐)	前体加工	[22~24]
类胡萝卜素		
羟化型	基因组合	[34]
新糖基化类胡萝卜素	基因组合	[35]
新类胡萝卜素途径	定向进化和基因组合	[19]
聚酮抗生素		
无环氧化作用的大环内酯	基因分裂	[8]
聚酮链延伸	模块组合	[14]
新大环内酯结构	模块组合	[15]
新大环内酯结构	前体加工	[26]
大环内酯环新修饰	基因分裂和加工	[25]
糖基化	基因组合	[11,12,36~40]
糖基化	定向进化	[21]
非核糖体肽		
细菌/真菌 NRP 杂合	模块组合	[13]
新非蛋白质氨基酸作为构件	基因分裂和前体加工	[27]

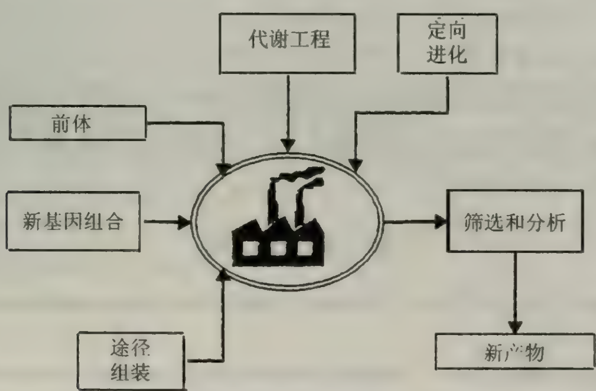


图 8.11　工程微生物细胞的途径

可以完全是来自一个机体的克隆或来自各种机体的基因组装成生物合成途径。例如，几个天然类胡萝卜素生物合成基因已经组合到非产生类胡萝卜素*E. coli*合成类胡萝卜素途径。微生物和植物的基因已组合到 *E. coli* 中产生紫杉醇生物合成的中间体——taxadiene。许多次级代谢产物的生物合成基因簇集在操纵子中，大大促进它们的克隆。在一个机体中建立一种新途径需要异源基因在宿主中表达，并且不产生有毒产物，一般优选 *E. coli* 作为宿主。

(2) 途径组装

产生特殊产物的途径可以几种方式进行操作。途径中的各种酶可以修饰。基因可以缺失或插入，不同途径的基因组合可以组装成新途径(组合生物合成)。该途径广泛用于构建新糖基的聚酮抗生素[9~10]。

(3) 模块酶

在一种途径中酶活性新的组合不能只是由组合整个基因来实现。多结构域的酶，具有特异活性的结构域可以交换。该途径已用于聚酮和非核糖肽(NRP)合成中模块酶的构建[13]。具有长链的新聚酮已通过在现有的 PKS 中插入额外的模块延伸聚酮合酶(PKS)获得。新的大环内酯也通过杂合模块系统合成。杂合的 NRP 通过转换细菌和真菌的 NRP 结构域获得。

(4) 定向进化

天然酶通过连续几轮基因复制，趋异，突变，重组和选择获得进化酶。进化酶通常保持它们的化学性质，但获得新的底物特异性。在生物合成途径中关键酶的进化可以产生新的化合物。收集代谢途径中各种酶的结构与功能资料将促进构建新途径。

基因组改组和染色体工程的新方法似乎成为制造和最佳化新途径的重要工具。

(5) 前体指导生物合成

新化合物可以通过新的构件移入生物分子中产生，这样的构件可以加入到机体生长的介质中，机体可以引入并代谢它们。该法已成功地用于产生新的多羟基链烷酸酯，聚酮抗生素和 NRP。

参 考 文 献

1 Pédelacq J D, Piltcn E, Liong E C, *et al*. Engineering soluble proteins for structural genomics. Natl Biotechnol, 2002, 20: 927 ~ 932

2 Yang J K, Park M S, Waldo G S, *et al*. Directed evolution approach to a structural genomics project: Rv2002from Mycobacterium tuberculosis. Proc Natl Acad Sci USA, 2003, 100: 455 ~ 460

3 Armstrong N, de Lencastre A, Gouaux E. A new protein folding screen: application to the ligand binding domains of a glutamate and kainate receptor and to lysozyme and carbonic anhydrase. Protein Sci, 1999, 8: 1475 ~ 1483

4 Sun L, Petrounia I P, Yagasaki M, *et al*. Expression and stabilization of galactose oxidase in *Escherichia coli* by

directed evolution. Protein Eng, 2001, 14: 699 ~ 704

5 Waldo G S, Standish B M, Berendzen J, *et al*. Rapid protein-folding assay using green fluorescent protein. Nat Biotechnol, 1999, 17: 691 ~ 695

6 Kim C A, Phillips M L, Kim W, *et al*. Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression. EMBO J, 2001, 20: 4173 ~ 4182

7 Yang J K, Park M, Waldo G S, *et al*. Crystallization and preliminary X-ray crystallographic analysis of the Rv2002 gene product from Mycobacterium tuberculosis, a β-ketoacyl carrier protein reductase homologue. Acta Crystallogr D, 2002, 58: 303 ~ 305

8 Pedelacq J D, Piltch E, Liong E C, *et al*. Engineering soluble proteins for structural genomics. Nat Biotechnol, 2002, 20: 927 ~ 932

9 Wurth C, Guimard N K, Hecht M H. Mutations that reduce aggregation of the Alzheimer's A beta 42 peptide: an unbiased search for the sequence determinants of A beta amyloidogenesis. J Mol Biol, 2002, 319: 1279 ~ 1290

10 Kawasake M, Inagaki F. Random PCR-based screening for soluble domains using green fluorescent protein. Biochem Biophys Res Commun, 2001, 280: 842 ~ 844

11 Waldo G S. Genetic screens and directed evolution for protein solubility. Current Opinion in Chemical Biology, 2003, 7: 33 ~ 38

12 Maxwell K L, Mittermaier A K, Forman-Kay J D, *et al*. A simple *in vivo* assay for increased protein solubility. Protein Sci, 1999, 8: 1908 ~ 1911

13 Sieber V, Martinez C A, Arnold F H. Libraries of hybrid proteins from distantly related sequences. Nat Biotechnol, 2001, 19: 456 ~ 500

14 Martin A, Sieber V, Schmid F X. *In vitro* selection of highly stabilized protein variants with optimized surface. J Mol Biol, 2001, 309: 717 ~ 726

15 Minard P, Scalley-Kim M, Watters A, *et al*. "loop entropy reduction" phage-display selection for folded amino acid sequences. Protein Sci, 2001, 10: 129 ~ 134

16 Auf der Maur A, Escher D, Barberis A. Antigen-independent selection of stable intracellular single-chain antibodies. FEBS Lett, 2001, 508: 407 ~ 412

17 Wigley W C, Stidham R D, Smith N M, *et al*. Protein solubility and folding monitored *in vivo* by structural complementation of a genetic marker protein. Nat Biotechnol, 2001, 19: 131 ~ 136

18 Cha H J, Srivastava R, Vakharia V N, *et al*. Green fluorescent protein as a noninvasive stress probe in resting *Escherichia coli* cells. Appl Environ Microbiol, 1999, 65: 409 ~ 414

19 Cortaxxo P, Cerveňansky C, Marin M, *et al*. Silent mutations affect *in vivo* protein folding in *Escherichia coli*. Biochem Biophys Res Comm, 2002, 293: 547 ~ 551

20 Lesley S A, Graziano J, Cho C Y, *et al*. Gene expression response to misfolded protein as a screen for soluble recombinant protein. Protein Eng, 2002, 15: 153 ~ 160

21 Knaust R K C, Nordlund P. Screening for soluble expression of recombinant proteins in a 96-well format. Anal Biochem, 2001, 297: 79 ~ 85

22 Peabody D S, Al-Bitar L. Isolation of viral coat protein mutants with altered assembly and aggregation properties. Nucleic Acids Research, 2001, 29: 113

23 Patnaik R, Louie S, Gavrilovic V, *et al*. Genome shuffling of lactobacillus for improved acid tolerance. Nat Biotechnol, 2002, 20: 707 ~ 712

24 Zhang Y X, Perry K, Vinct V A, *et al*. Genome shuffling leads to rapid phenotypic improvement in bacteria. Na-

ture, 2002, 415: 644 ~ 646

25 张惠展. 途径工程. 北京:中国轻工业出版社,2002

26 Lassner M, Bedbrook J. Directed molecular evolution in plant improvement. Current Opinion in Plant Biology, 2001, 4: 152 ~ 156

27 Matzumura I, Wallingford J B, Surana N K, *et al*. Directed evolution of the surface chemistry of the reporter enzyme β-glucuronidase. Nat Bioteehnol, 1999, 17: 696 ~ 701

28 Rowe L A, Geddie M L, Alexander O B, *et al*. A comparison of directed evolution approaches using the β-glucuronidase model system. J Mol Biol, 2003, 332: 851 ~ 860

29 Seffernick J L, Johnson G, Sadowsky M J, *et al*. Substrate specificity of atrazine chlorohydrolase and atrazine-catabolizing bacterria. Appl Environ Microbiol, 2000, 66: 4247 ~ 4252

30 Honee G, van der Salm T, Visser B. Nucleotide sequence of crystal protein gene isolated from *B. thuringiensis* subspecies entomocidus 60.5 coding for a toxin highly active against Spodoptera species. Nucleic Acids Research, 1988, 16: 6240

31 Sandmann G, Albrecht M, Schnurr G, *et al*. The biotechnological potential and design of novel carotenoids by gene combination in *Escherichia coli*. TIBTECH, 1999, 17: 233 ~ 237

32 Schmidt-Dannert C, Umeno D, Arnold F H. Molecular breeding of carotenoid biosynthetic pathways. Nat Biotechnol, 2000, 18: 750 ~ 753

33 Albrecht M, Takaichi S, Steiger S, *et al*. Novel hydroxycarotenoids with improved antioxidative properties produced by gene combination in *Escherichia coli*. Nat Biotechnol, 2000, 18: 843 ~ 846

34 Ye X, Al-Babili S, Kloti A, *et al*. Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science, 2000, 287: 303 ~ 305

35 May O, Nguyen P T, Arnold F H. Inverting enantioselectivity by directed evolution of hydatoinase for improved Production of L-methionine. Nat Biotechnol, 2000, 18: 317 ~ 320

36 Soong N W, Nomura L, Pekrun K, *et al*. Molecular breeding of viruses. Nature genetics, 2000, 25: 436 ~ 439

37 Powell S K, Kaloss M A, Pinkstaff A, *et al*. Breeding of retroviruses by DNA shuffling for improved stability and processing yields. Nat Biotechnol, 2000, 18: 1279 ~ 1282

38 de Boer A L, Schmidt-Dannert C. Recent efforts in engineering microbial cells to produce new chemical compounds. Current Opinion in Chemical Biology, 2003, 7: 273 ~ 278

订购 2004.11.19
收到期 2004.12.8.
来 源 科学出版社
书 价 28.60元
单据号 9964421
开票日期 2004.12.2

中科院植物所图书馆
S0003774

58.17435
552
000028131
2004
书名 进化生物技术:酶定向分子进化

借者姓名	借出日期	还书日期
梁[illegible]	05.3.24	

58.17435
552
000028131

(Q-1360.0101)

现代生物技术前沿丛书

书　名	书号　ISBN	定价
基因免疫的原理和方法	7-03-012588-6	38元
RNAi：基因沉默指南（影印版）	7-03-012685-8	65元
从基因到基因组——DNA技术概念和应用（影印版）	7-03-012477-4	46元
结构生物学与药学研究	7-03-011688-7	48元
生物芯片分析（影印版）	7-03-012247-X	80元
DNA芯片和基因表达（影印版）	7-03-012248-8	29元
计算分子生物学导论（翻译版）	7-03-011493-0	36元
植物生物技术导论（影印版）	7-03-012799-4	72元
蛋白质化学与蛋白质组学	7-03-012401-4	75元
进化生物技术——酶定向分子进化	7-03-012639-4	38元

ISBN 7-03-012639-4
9 787030 126399 >

生命科学编辑部
联系电话：010-64012501
http://www.lifescience.com.cn
e-mail:info@lifescience.com.cn

ISBN 7-03-012639-4
定 价：38.00 元